DOCUMENT RESUME

ED 313 780 EA 021 443

TITLE	Policy Recommendation on Performance-Based Accreditation. A Report Prepared for the Commissioner of Education and the State Board of Education of the Nebraska Department of Education.
INSTITUTION	Nebraska Univ., Lincoln. Bureau of Educational Research and Field Services.
SPONS AGENCY	Nebraska State Dept. of Education, Lincoln. Div. of Vocational Education.
PUB DATE	3 Jan 89
NOTE	101p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Academic Achievement; *Accreditation (Institutions); Change Strategies; *Curriculum; *Educational Assessment; Educational Change; *Educational Policy; *Educational Quality; Elementary Secondary Education; Excellence in Education; *Testing
IDENTIFIERS	*Nebraska

ABSTRACT

The content of this report, commissioned by the State Board of Education, examines the current input indicators contained within Nebraska Statutes. Recommendations for ways in which process performance indicators suggested by the Accreditation Advisory Committee may be appropriately evaluated and measured are provided. The common curriculum of Nebraska's school system is identified and recommendations are made for appropriate testing strategies. A review is made of the Comprehensive School Evaluation Process currently required in state and in North Central accreditation. The majority of the report is contained in seven appendices: (1) charge from the State Board of Education and makeup of Task Force; (2) catalogue of indicators; (3) textbook usage in elementary schools of educational service Unit #3; (4) partial summary of course offerings in Nebraska secondary schools; (5) testing options and costs, National Assessment of Education Program, achievement batteries, and subject area tests; and (6) North Central Accreditation Process. (32 references) (SI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED313780

BULLETIN

POLICY RECOMMENDATION ON PERFORMANCE-BASED ACCREDITATION

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ward Lyfoute

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Bureau of Educational Research and Field Services
Department of Educational Administration
University of Nebraska
Lincoln, Nebraska

EA 021 443

SA

POLICY RECOMMENDATION ON

PERFORMANCE-BASED

ACCREDITATION

A REPORT PREPARED FOR THE
COMMISSIONER OF EDUCATION
AND
THE STATE BOARD OF EDUCATION OF THE
NEBRASKA DEPARTMENT OF EDUCATION
LINCOLN, NEBRASKA

Janaury 3, 1989

The Bureau of Research, Service and Policy Studies
Department of Education Administration
University of Nebraska
Lincoln, Nebraska

University of Nebraska Lincoln

Department of Education Administration
1204 Seaton Hall
Lincoln, NE 68588-0638
Telephone: (402) 472-3726

Commissioner Joe E. Lutjeharms
301 Centennial Mall South
Lincoln, Nebraska 68509-4987

Dear Commissioner Lutjeharms:

The content of this report, as commissioned by the State Board of Education, addresses: (1) the examination of the current input indicators which are contained within Nebraska Statutes; (2) recommendations for ways in which process performance indicators suggested by the Accreditation Advisory Committee may be appropriately evaluated and measured; (3) identification of the common curriculum of Nebraska's school system and recommendations for appropriate testing strategies; and (4) a review of the Comprehensive School Evaluation Process currently required in State and in North Central accreditation.

The Task Force that undertook this responsibility made every effort possible within the time available to it, to search through the literature, review what had been done in other states regarding accreditation, review the testimony presented at hearings held by the Nebraska State Accreditation Committee, study previous reports made by committees regarding accreditation in Nebraska, and take into consideration suggestions made by advisory committees of administrators and classroom teachers. While the Task Force reviewed all materials it had brought together with advisory groups of administrators and teachers, it was natural that from such groups there were various views, interests, and opinions. With the diversity of opinions, it was not possible to always reach unanimous agreement, even though all points and options were taken into account. This report, which was the responsibility of the Task Force, reflects the current thinking of specialists on the subjects related to accreditation while taking into account the views of Nebraska educators.

The monumental effort of the Task Force is distilled in the pages of this report which will provide the State Board of Education, as well as other interested parties, state of the art information regarding the complex issue of school accreditation. Hopefully, this report will provide a database that will be of help to policy makers and ultimately enhance the quality of education for the children and youth in the State of Nebraska.

Respectfully,

Ward Sybouts

Ward Sybouts, Director
Bureau of Education Research,
Service and Policy Studies

TABLE OF CONTENTS

PAGE

POLICY RECOMMENDATION ON PERFORMANCE-BASED ACCREDITATION . . . 1

Report to the Nebraska Board of Education. 1
Definitions and Explanations of Terminology. 6
Comments on Methodology. 9
Responses to Work Assignment 11

RECOMMENDATIONS. 17

APPENDIX A - The Charge from the State Board of Education and Makeup of Task Force. 23

APPENDIX B - A Catalogue of Indicators 29

APPENDIX C - Textbook Usage in Elementary Schools of Educational Service Unit #3 35

APPENDIX D - Partial Summary of Course Offerings in Nebraska Secondary Schools. 39

APPENDIX E - Testing Options and Costs 43
National Assessment of Education Program. 52
Achievement Batteries 65
Subject Area Tests. 78

APPENDIX F - North Central Accreditation Process 91

APPENDIX G - Selected References 97

POLICY RECOMMENDATION ON PERFORMANCE-BASED ACCREDITATION

Report to the Nebraska State Board of Education

This report, prepared by the University of Nebraska Bureau of Education Research, Service and Policy Studies, is in response to action of the State Board of Education at its September 9, 1988 meeting. The specific requests of the State Board as well as the composition of the Task Force responsible for this report are found in Appendix A.

Although the Task Force accepts its charge from the State Board of Education as confined to the specific tasks assigned it, the response to that charge and the results of the Task Force's work need to be set in the context of larger issues surrounding the accreditation of Nebraska's schools. The Task Force suggests that the accreditation system should provide a broad framework which includes provisions to address accreditation, accountability, evaluation, and school improvement. The performance-based accreditation system should provide clearly defined incentives to support local school district efforts to improve the quality of education available for students in Nebraska. Decisions regarding final answers are intimately related to policy issues and are outside the scope or responsibility of the Task Force's work. Nevertheless, in order to sharpen the focus for policy makers, the following outline is a necessary introduction to the Task Force's findings.

I. The quest for change from the current accreditation and approval procedure is part of a national trend demanding more accountability. Citizens want to know in concrete terms what their tax dollars are buying. Citizens are concerned about student performance, school results, and progress toward school improvement.

A. The intense national activity in this arena provides Nebraska the advantage of profiting from the work and experience of other states. At the same time, there are good reasons for Nebraska to be wary of blindly following the lead of other states. Nebraska's specific needs and conditions must be addressed in Nebraska policy considerations.

1. Initiatives in other states are in initial and evolving stages and differ one from another.[1] Many are less than successful at this point in the sense that they have created as many problems as they have attempted to solve, despite expenditures of relatively large resources in time, effort, and money. States must insure that adequate capacity, such as computer systems, staff, and training, is in place before implementing an expanded accountability system. Given the present Nebraska Department of Education fiscal condition, the accountability system needs to be reduced rather than expanded and should become more focused in an integrated manner to supply meaningful statewide data (see Appendix E, Item 5, p. 51).

2. Nebraska is not a typical state. We are at the extreme on the spectrum of state (as opposed to local) fiscal support of public schools and equally as extreme on the spectrum in terms of traditional commitment to the concept of local control of school districts.

3. Advice from educators, citizens, and policy makers agrees that plans should be tailor made to each state and that such action should be carefully timed rather than precipitous.[2]

---

[1]See, for example, "Creating Responsible and Responsive Accountability Systems: Report of the OERI State Accountability Study Group" (Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education, September, 1988). This report, the result of a year-long study, collected extensive information from all fifty states and dealt in-depth with the experiences of ten states.

[2]Ibid.

B. Nebraska policy makers believe that little is known about the quality of education in Nebraska and that a significant accountability gap exists.[3] However, many educators hold the point of view that a great deal is known in each district about the quality of education in that district but the data are not in a format that can be readily aggregated to provide an easily understood statewide picture. The combination of strong local control and wide diversity of districts in size and in program exacerbates the difficulty.

C. On some national measures used to compare states, Nebraska's educational system ranks among the nation's strongest.[4] For some, the high ranking of Nebraska schools brings into question the need for reform. However, because of the questionable validity of the measures reported, the Council of Chief State School Officers has initiated a well financed and carefully planned effort to develop a better system for making state to state comparisons.[5] As the Council's project matures, Nebraska will get a more accurate picture of its relative standing, without reference to its own efforts to develop or use more sophisticated measures.

II. Although the accountability questions are disarmingly simple, developing answers is a complex problem and raises a number of difficult issues.

A. Clarity of purpose is fundamental. The data and processes necessary to provide oversight by the State Department of Education may differ from the data and the processes required to ensure improvement. Systems that serve both purposes must find the appropriate balance between the two functions. Balance must also

---

[3]J. P. Sher, Class Dismissed: Examining Nebraska's Rural Education Debate (Lincoln: Nebraska Rural Community Schools Association, 1988). See also, Evaluation and Assessment Processes in Nebraska Public Schools, Interim Study LR 181 (Lincoln: Nebraska Legislative Council, February 1988).

[4]The Condition of Education (Washington, D.C.: National Center for Education Statistics, U.S. Department of Education). See also, Results in Education, The Governors' 1991 Report on Education (Washington, D.C.: National Governors' Association, 1988).

[5]Council of Chief State Officers, Education, Evaluation, and Assessment in the United States (Washington, D.C.: The Council, 1984).

be identified and more clearly understood between the responsibility of the state as compared to the local district's responsibility for both oversight and improvement (see Appendix E, Item 1, p. 51).

Although clearly outside the charge given to the Task Force, consideration of purpose and function demands an answer to questions about what is to be done with the data that are gathered. Questions, such as whether districts not meeting established standards are to be chastised or are to be given additional resources; whether providing more resources becomes a disincentive for improvement; and whether high achieving districts are to be rewarded and, if so, how, all speak to the heart of the question of purpose.

B. Deciding who should be held responsible for which outcomes involves matching accountability instruments with the responsible person or institution. Such matching requires understanding the purposes and the limits of different accountability measures, not only by who uses them, but more importantly, by those who interpret results. For example, student test scores are seldom valid proxies for teacher performance.[6]

C. State mandates, especially with regard to testing programs, tend to drive curriculum decisions and thereby threaten local control.[7] However, it is necessary to establish a state standard as a base in order to ensure that students in Nebraska can perform at the State's expectation. State-mandated accountability systems are powerful policy tools, but they may also provide some unintended effects that seriously constrain instruction. State-controlled systems often result in narrowing of curriculum without necessarily ensuring qualitative improvement.

A related issue requires that policy makers, citizens, and educators be informed regarding expected performance levels.

---

[6]Ronald Edmmonds, "School Effects and Teacher Effects," Social Policy, 15 (Fall 1984).

[7]Harold Berlak, "Testing in a Democracy," Educational Leadership, 43 (October 1985). See also, Noe Medina and D. Month Neill, Fallout from the Testing Explosion (Cambridge, Massachusetts: National Center for Fair and Open Testing, 1988).

D. Available commerical standardized tests are imperfect instruments for measuring academic achievement.[8] Market available instruments do not measure the full range of academic achievement across all curriculum areas. Currently available instruments rarely measure higher-order skills or match what is being taught in a specific school building. Reliance on multiple-choice or short answers is simply not adequate as a single measure of student performance. In fact, the use of a single instrument tends to standardize and narrow the curriculum.

E. Given the significant differences among Nebraska school districts in characteristics and resources, it is extremely difficult to find a fair basis for comparing results using a single standardized norm-referenced test or a variety of commercially available standardized norm-referenced tests.

III. The dilemmas described above will not be solved by simply rewriting Nebraska Department of Education Rules 14 and 15. Members of the Task force see establishing a new basis for accreditation as only the beginning point in a longer-term process which holds great promise for improving education in Nebraska. Essentially, the process involves establishing a more systematic, specifically designed management information plan. The management plan should be based on the purposes for collecting data, decisions about which data support those purposes, and judgments about how the data can be most effectively and efficiently gathered, interpreted, and distributed, and how this information can best inform policy decisions. The system must also be flexible to accommodate the rapidly expanding body of research in the testing movement and the growing interest in accountability indicators and their use. The Task Force suggests there is a critical need for an information management system which will integrate the above-suggested processes and which will enable continous and in-depth examination and updating of education data in Nebraska. The task of establishing a comprehensive data base offers an opportunity for the evolution of consensus on questions which are divisive and the development of trust among all participants.

---

[8] James H. Lytle, "The Zest for Tests," The School Administrator, 45, #11 (December 1988).

IV. Concerns brought to the attention of the Task Force by the Commissioner's appointed advisory group, teachers, and individual Nebraska educators suggest multiple agendas and differentiated concerns among those involved. Past controversy about school district reorganization and current concerns about distribution of state funds to school districts contribute to uneasiness and suspicion about motives in any suggested changes. To the extent that apprehension and mistrust stem from misunderstanding, it becomes critically important to gain common acceptance of precise definitions of key terminology. Accordingly, this report begins with a set of definitions.

Definitions and Explanations of Terminology

Although the following definitions generally agree with commonly accepted usage, it is important to note that these definitions of accreditation and approval do not match the way those terms have been historically used in Nebraska. The purpose of presenting definitions at this point is to emphasize that these are the specific meanings for the terms as used in this report.

Accreditation. The process by which the State Department of Education determines that a school district meets the minimum criteria required by the state of Nebraska. Accreditation automatically signifies approval. Non-public schools may seek accreditation by application.

Approval.[9] The process by which the State Department

---

[9]A definition of approval is included here only because it is terminology appearing in the statute. The Task Force believes "recognized" is the more appropriate word to fit this definition on the basis that a school building may be allowed to operate, as in the case of Home Schools, without the State Department necessarily approving of the operation.

determines that a non-public school can be allowed to operate in Nebraska.

Recognized. Any public school authorized by law to exist and operate in Nebraska.

Accountability. Responsibility of program staff and administrators to provide evidence to funders and superiors of conformity to program specifications and fiscal requirements.[10]

Performance accountability system. A systematically collected set of data or indicators that provide information about how school districts are performing. The word system implies that no matter how valid and reliable any single indicator may be, it cannot be interpreted independently of other indicators or data that reflect the total educational institution.

Assessment. The process of gathering data and putting it into an interpretable form. Assessment is a preliminary step to making an evaluative judgment or a decision about a program.[11]

Evaluation. A process for determining how well performance-based outcomes are reached; connoting a qualitative judgment. "Evaluation is the determination of a thing's value. In education, it is the determination of the quality, effectiveness, or value of a program, product, project, process, objective, or curriculum.[12]

---

[10] Evaluation: A Systematic Approach (Beverly Hills, California: Sage, 1979).

[11] S. Anderson, S. Ball, R. Murphy, and Associates, Encyclopedia of Educational Evaluation (San Francisco: Jossey-Bass, 1975).

[12] Blaine R. Worthen and James R. Sanders, Educational Evaluation (New York: Longman, 1987).

Indicators. Data or information that reveal something about the health or performance of the educational institution.

For purposes of this report, it is useful to separate indicators and data into several categories. The rationale for establishing categories is to better understand the relationships among the various indicators and data. For example, if external contextual data impact the quality of education but are outside the control of school officials, it is obviously unfair to compare school buildings without taking those data into account. Similarly, if comparisons must be made, then a level ground as a basis for comparison can be established by grouping school buildings whose programs are provided within similar contexts.

External contextual data. Variables which describe the environment in which the educational program is provided. They describe that part of the environment over which the school district has no control but which, nevertheless, impact educational results.

Resource data. Those contextual data which describe the wherewithal which enables an educational institution to carry out its mission.

In-put indicators. Those measures of effort and/or the configuration of resources an educational institution chooses to apply to its mission.

Process indicators. Those procedures and/or routines used by an educational institution in carrying outs its mission.

Outcome indicators. The results of an educational program.

Norm-referenced tests. The scores from norm-referenced tests give meaning by relating performance to the overall distribution of tests scores from a norm group. The norm group can be various referent groups, but often is a nationally representative sample of examinees.

Criterion-referenced tests. The scores from criterion-referenced tests give meaning by relating performance on test items to pre-established criteria for acceptable performance. Usually several items are developed to measure a single objective and scores are reported in terms of mastery/non-mastery of test objectives.

Curriculum. Those activities, courses, and materials by which instruction is delivered and for which the school district assumes responsibility.

Common curriculum. The part of the curriculum which is generic and in which all students of a given category are expected to participate.

Comments on Methodology

The task assigned the Task Force involved an examination of current indicators used in the Nebraska accountability system as well as looking at some suggested additional indicators. In approaching this assignment, the Task Force

developed a comprehensive array of potential indicators and relevant data items. With regard to each item, the following kinds of questions were posed.

1. Can the indicator be measured?

2. Does research support the indicator as a valid measure of quality education?

3. Are the collection and use of the data or indicator reasonably cost effective in terms of improving education in Nebraska?

4. How successfully has the indicator been used in other states?

An extensive list of indicators was examined to determine if the indicators met the criteria of usefulness and cost effectiveness in relation to qualitative impact on education. Indeed, the critical measure of an accountability system's usefulness is whether it provides information that can be easily understood by a broad audience and whether the burden of collecting, analyzing, and reporting is within the existing time and cost constraints. All of these requirements suggest the careful selection of a few key indicators.

The list of indicators is found in Appendix B. Each indicator is categorized with regard to collection point (building or district level), utility with respect to approval, accreditation or evaluation, and to whom it is to be reported. Those data or indicators marked with an asterisk in Appendix B are recommended by the Task Force for inclusion in the state accountability system.

Members of the Task Force met repeatedly with an advisory committee of practitioners during the development of this report. Also, a representative group of classroom teachers selected by the Nebraska State Education Association was provided an opportunity to react to the report. Although members of the Task Force attended carefully to reactions of advisory groups, the Task Force takes full and total responsibility for the contents of this report.

Responses to Work Assignment

The Task Force reports the following findings in response to the requests of the State Board of Education.

1. The Task Force examined the various indicators in the accreditation and approval system that arise from Nebraska statutes and regulations. The requirements range from very general to quite specific and cover personnel, program, facilities, and materials. The requirements are not cohesive in the sense that there is a comprehensive statute or collection of statutes that relate directly to a set of goals for education in Nebraska. Although all the statutes are intended to establish or maintain qualitative standards, separately they speak to a variety of specific objectives. The requirement for a specific number of hours of instruction, for example, speaks to a currently popular public opinion that time in school is related to quality education.

Some statutes exist to provide flexibility of interpretation of previous enactments in order to accommodate the existing

variance in capability or resources among Nebraska schools. Allowing for the delivery of course content requirements on an every-other-year or an off-site basis is an example of variance permitted.

It is the position of the Task Force that time and energy spent in checking on the enforcement of a set of statutes do not contribute significantly to the qualitative improvement of schooling in Nebraska. This is not to say that the statutes should not be enforced, but is to point out that a detailed enforcement process does not necessarily provide information useful either for improving instruction or for reporting to the public about the state of education in Nebraska.

We believe a simple sign-off procedure coupled with random audits would better serve the enforcement purpose.

2. The task Force examined the various process indicators suggested in early reports completed in Nebraska and other states for inclusion in an accreditation system. Without in any sense diminishing the importance of process, the Task Force questions the usefulness of process indicators as part of the accreditation system. It may well be that the most productive way to provide for local control, initiative, and creativity is through allowing flexibility in process. Such thinking follows the philosophy that it is the function of the state to set standards and goals and to judge if they are met, while procedures for accomplishing those goals and standards be best left to local discretion.

Limiting the use of process indicators helps in the difficult task of reducing the data gathering and reporting burden. It also strikes an important compromise and lessens the tensions related to shifts from local to state control.

3. The Task Force took a variety of approaches in attempting to identify the common curriculum of Nebraska schools. In a general sense, the curriculum is already defined by statute which outlines the mission of schooling and the areas of learning which must be covered. An analysis of elementary textbooks in use in Nebraska suggests sufficient variety to rule out this avenue to identify common curriculum.[13]

The State Department of Education collects numerous reports regarding curricular offerings and instructional personnel qualifications for courses being taught at the secondary school level.[14] However, an examination of the last two years of such reports made by school districts to the State Department of Education fails to identify a common curriculum in the high schools of Nebraska. The reports do show that a variety of courses is provided. Attempts to interpret the reports also demonstrate the difficulty of knowing what is taught on the basis of the title of the course.

---

[13]A summary of textbook usage in Educational Service Unit #3 is presented in Appendix C.

[14]A partial summary of findings related to secondary school curriculum reports is found in Appendix D.

At the elementary level, the State Department publication, "Planning the Weekly Schedule in Elementary Schools" comes the closest to describing a common elementary curriculum. Determining the extent to which this guideline is followed was not attempted by the Task Force, and there are no reports to the state from school districts regarding the elementary curriculum.

Prior to determining the appropriateness of performance-based measures, a more careful definition of what is to be learned must be determined. A collection of outlines of school district testing programs now in existence indicates considerable standardized testing of students is already in progress at the district level. A state testing program, as compared to local programs, should await a clear definition of state expectations and must also reflect the state's purpose for testing. Although it has been suggested that the National Assessment of Education Progress (NAEP) tests might be useful in this context, the NAEP test batteries are not available to states at the present time. A subreport (see Appendix E) provides an analysis of the various testing options, the consequences of employing each, and an estimate of the costs of implementation.

At present, until state outcomes for learning have been established, reporting locally determined outcomes could serve state policy makers in regard to how Nebraska students are performing at this local district level.

4. The Task Force examined the North Central accreditation process. Appendix F provides a summary of the indicators

used in the latest version of the North Central accreditation process. Keeping in mind that the North Central Association is a voluntary regional organization of schools and college, and that its original purpose is to accredit high schools and colleges with reference to standardization of college entrance requirements, it is the opinion of the Task Force that North Central accreditation should not be a part of the Nebraska accreditation system. This is not to impugn the North Central accreditation process. The North Central process is outside the jurisdiction of Nebraska school officials and may not necessarily serve the purposes of an evolving Nebraska accreditation system.

RECOMMENDATIONS

Beyond the findings of this report and as a result of its investigations, the Task Force feels reponsible to make some initial recommendations dealing with the larger question of a state accountability system. The recommendations are influenced by several philosophical assumptions.

The purposes of enforcing minimum standards, of providing for the improvement of instruction, of supporting equal educational opportunities for all students, and of encouraging change to restructure schooling cannot all be served by the same single mechanism. However, a broadly conceived state management information system encompassing accreditation, providing for evaluations, and encouraging innovation may accomplish multiple purposes. Purposes must be clear for each segment of the system. Confusion related to terminology must be eliminated.

If change is a prime purpose, the imposition of more tightly coupled controls and of standardization and centralization is counter productive. The several important national attempts at restructuring schools start with a waiver of existing controls. Nebraska school districts already enjoy a great degree of self-determination which should be protected to the extent possible.

Present statutes and current rules and regulations are subject to change. Such changes should reduce rather than increase reporting requirements.

As the reader considers the following recommendations,

it is important to have studied the text of this report and to have become familiar with the content in the various appendices. Some of the recommendations which follow summarize points already made elsewhere in the report.

1. The statutes, rules, and regulations which deal with educational improvement should be thoroughly reviewed and recodified to provide a more cohesive and simplified set of requirements related to the purpose for each. The appropriate roles of the Legislature, the State Board of Education, local boards of education, and State Department of Education staff with regard to assessment and evaluation should be more carefully delineated.

2. The current reporting forms required by the state of local school districts should be examined to determine which can be eliminated. Further, consideration should be given to having the chairperson of the local board of education certify annually, in a single statement, that the school district is in compliance with rules and regulations. The superintendent of schools will provide the board with such assurance.

3. The information management system within the State Department of Education should be reorganized to collect and analyze information that will contribute to an annual report to the public on the state of education in Nebraska. The report should be, at minimum, in a form that will provide longitudinal data showing trends in meeting state goals. More resources in personnel, hardware, and dollars should be provided than currently exist in the State Department of Education in order that

the public may be adequately informed of the condition of education in Nebraska. Furthermore, a statewide testing program would further exacerbate the lack of resources.

4. The Nebraska State Department of Education should develop, with the assistance of practitioners, learning goals for each approved K-12 curriculum area. As learning goals are stated, appropriate state measures of student performance should be specified.

5. As an initial step in reporting student performance, local school districts should report to the State Department of Education the results of district-selected standardized tests in all curricular areas tested.[15]

As the state defined curriculum areas are specified and measures developed to determine student performance, these

---

[15]There has been an increase in recent years of school-by-school comparisons on the basis of norm-referenced tests. This practice apparently stems from the rationale that the public has a right to know aggregated test results and that publicizing aggregated scores from school buildings will motivate better student performance. The Task Force rejects this reasoning for sound educational reasons. Comparison of schools is a misuse of tests because they were not developed for that purpose. More importantly, low scores are positively correlated with the incidence of social disadvantagement, poverty, ethnic minority status, and other environmental conditions. It is, therefore, very difficult to make fair school-to-school comparisons on the basis of standardized test scores.

It is interesting to note that statute 79-328 of Nebraska law, providing for testing of students in private, denominational, and parochial schools includes the following language: "The results of such testing . . . shall not be used to measure, compare, or evaluate the competency of students at such schools." In like manner, it is the understanding of the Task Force that there is a similar prohibition of the use of NAEP tests to compare individual schools.

results should also be reported to the citizens of the state and the Legislature.

6. School districts should be required to make an annual report to their publics,[16] as prescribed by the State Board of Education. The report would include at least the following: student performance, district demographics, and financial information.

7. North Central Association accreditation should not be considered a part of the state accreditation system. Continuation of North Central Association membership is, however, encouraged as an additional volunteer mechanism for school improvement.

8. A new state evaluation process should be developed to include on-site audit visitations more often than the present seven-year cycle. The audit visitation should be conducted in order to examine evidence that statutes are being met and process indicators are being appropriately applied, and to determine if efforts are being made to address any previously identified problems and/or to improve upon district performance.

9. The state of Nebraska should develop a statewide accreditation system which provides the vehicle to collect data in a systematic manner over time to determine the condition and status of education at the local district level and at the state

---

[16]It is assumed that almost all school districts already make an annual report to their board of education. The rationale for this recommendation is to promote some standardization of the contents of such reports so that together they help provide information of a statewide nature.

level. Data and indicators which impact education should be collected and reported on an annual basis. The system should have the flexibility to respond to the develompent and availability of more sophisticated tests of student performance and to the growing body of research about the linkages between indicators and student results.

10. The State Department of Education should examine the advantages of statewide assessments on a single or a few items of performances in a given school year, with appropriate rotation of items to be assessed from year to year. Such procedure provides flexibility to respond more readily to schooling problems arising from changes in the environment and also provides for more in-depth study of any one item at the time it is assessed. Determining overall outcomes then will require linking the parts in a systematic manner.

11. All information coming to the State Department from school districts should come to the management information department. This department should become the focal point for both gathering and disseminating information and should be the control point for the integration of all data bases under the Commissioner's jurisdiction.

12. Policy decisions should be made to clarify the meaning of core curriculum as used in the statutes, to translate and expand the statutory mission statement into goals and objectives, and to give guidance to local school districts regarding breadth

of curriculum and delineating the degree of school district control over sequencing of curriculum.

13. The State Board of Education and the Nebraska Department of Education should appoint advisory committees of practitioners and other appropriate persons to provide input regarding implementation of the recommendations of this report.

APPENDIX A

THE CHARGE FROM THE STATE BOARD OF EDUCATION AND MAKEUP OF TASK FORCE

THE CHARGE FROM THE STATE BOARD OF EDUCATION AND MAKEUP OF TASK FORCE

The report of the Task Force responds to items 3, 4, 5, and 6 of Commissioner Lutjeharms' report to the State Board of Education, a copy of which is attached.

The University of Nebraska Bureau of Education Research, Service and Policy Studies, formed a Task Force to prepare a report in response to the Commissioner's recommendations. Members of the Task Force are:

Dr. Robert Stalcup, University of Nebraska-Lincoln
Dr. Patricia O'Connell, University of Nebraska, Omaha
Dr. Barbara Plake, University of Nebraska-Lincoln
Dr. Robert Brown, University of Nebraska-Lincoln
Dr. Steven Wise, University of Nebraska-Lincoln
John Prasch, University of Nebraska-Lincoln
Dr. Ken Hansen, Omaha Westside District 66

The Administrator Advisory Committee is composed of the following persons:

Dr. Gary Bennett, Assistant Superintendent
Omaha Public Schools

Dr. Eleanor Francke, Program Coordinator
Lincoln Public Schools

Dr. Jim Davis, Assistant Superintendent
Bellevue Public Schools

Leon Dappen, Assistant Superintendent
Papillion-Lavista Public Schools

James Havelka, Superintendent
Rising City Public Schools

Dr. Doug Christensen, Superintendent
North Platte Public Schools

Dr. Roger Breed, Assistant Superintendent
Elkhorn Public Schools

Dr. Tom Nenneman, Assistant Superintendent
Millard Public Schools

The Teacher Advisory Committee consists of the following persons:

Carol Beaty
Waverly Public Schools

Carolyn Grice
Omaha Public Schools

Judy Harris
Beatrice Public Schools

Wes Henn
North Platte Public Schools

Mary Jo Huelle
Gering Public Schools

Pam Murphy
Fremont Public Schools

Ann Palmesano
Omaha Public Schools

Cheryl Rains
Wilbur Public Schools

Lee Ransey
Minden Public Schools

Bob Rogge
Lincoln Public Schools

Ruth Ann Wylie
Lincoln Public Schools

NEBRASKA DEPARTMENT OF EDUCATION
301 Centennial Mall South Box 94987
Lincoln, Nebraska 68509-4987 Telephone: (402) 471-2295

September 6, 1988

TO: State Board of Education

FROM: Joe E. Lutjeharms, Commissioner of Education

SUBJECT: Policy Recommendation on Performance-Based Accreditation

Introduction

The State Accreditation Committee is to be commended for the fine job they had done in formulating a plan which reponded to the information gathered from the survey and the hearings conducted in the state. Many of the suggestions made will be incorporated into a new state accountability system.

It is the responsibility of the State Board of Education to adopt the policies which will set the parameters for the development of a quality and performance accreditation system. The policies must promote educational equity and satisfy the compelling interest of the State in the education of its youth. The policies must also provide for State monitoring and oversight which assure educational quality control and legal operation of all schools. The following recommended policies meet these requirements and provide the direction for the development of a new quality and performance-based accreditation rule.

Recommendations

It is recommended that the State Board of Education adopt the following policies and procedural guidelines for the development of a quality and performance-based accreditation system for Nebraska public schools.

1. The classification and legal operation of all public schools will be based on compliance with one common set of requirements which will be called accreditation.

2. The schools currently classified as Approved would be allowed a reasonable period of time (at least two years) to comply with the new accreditation standards.

It would appear that one set of standards and one classification level for all public schools would eliminate the inequity implicit in a multilevel system. All students in Nebraska public schools are entitled to a uniform base of quality as established by the State Board of Education.

3. The current input indicators which are contained within Nebraska Statutes shall be examined by the Policy Analysis Center at the University of Nebraska and the Center shall make its findings and recommendations known to the State Board within four months.

4. The Policy Analysis Center at the University of Nebraska shall develop and recommend ways in which process performance indicators suggested by the Accreditation Advisory Committee may be appropriately evaluated and measured.

5. The Policy Analysis Center shall identify the common curriculum of Nebraska's school system and recommend appropriate testing strategies in order to measure performance within the common curriculum of Nebraska's public education system. Such findings and recommendations shall be reported to the State Board within four months.

6. The Policy Analysis Center shall review the Comprehensive School Evaluation Process currently required in State and in North Central Association accreditation and report its findings and recommendations to the State Board within four months.

7. The Double A (AA) classification of schools shall be eliminated from the accreditation section activity and shall be included as a part of a Nebraska School Recognition Program. Such a program would be a voluntary program established to provide incentives for school improvement and to recognize educational excellence in schools of any size. A State-level system organized to promote excellence would be more appropriately suited to Nebraska's unique demographics and societal aspirations.

APPENDIX B

A CATALOGUE OF INDICATORS

Indicator	Wh Coll
External Contextual Data	
Community Size	Dist
Community Population Distribution	Dist
Community Description (Urban-Rural-Suburban)	Dist
% Handicapped Students	Buil
% Minority Students	Buil
% Low Income Students	Buil
% Non-Public Students	Dist
% Mobility	Dist
Resource Data	
Community Wealth	Dist
Gross District Product[1]	Dist
Assessed Property Value	Dist
Assessed Value per Student	Dist
% Students in Preschool	Buil
Social Services Available[2]	Dist

*Items marked with an asterisk are recommended by accreditation and evaluation system.

[1]This item suggests the development of a measure o gross national product, which includes productivity a

[2]This item refers to the availability of social se It includes governmental agencies such as juvenile cou Service, and youth-serving agencies such as YMCA and

INDICATORS

ere ected	Where Reported	Utility	Recommen-dation
rict	State	Evaluation	*
rict	Local	Evaluation	
rict	State	Evaluation	*
ding	State & Local	Evaluation	*
ding	State & Local	Evaluation	*
ding	State & Local	Evaluation	*
rict	State & Local		*
rict	State & Local	Evaluation	
rict	State	Evaluation	
rict	State & Local	Evaluation	
rict	State	Evaluation	*
rict	State & Local	Evaluation	*
ding	Local	Evaluation	*
rict	Local	Evaluation	

the Task Force for inclusion in a comprehensive

f school district wealth patterned after the
s well as other measures of wealth.

rvices of the kind that support at-risk children.
rt systems, volunteer agencies such as Family
YWCA.

Indicator	Wh Coll
Resource Data (continued)	
Educational Services Available	Dist
Formal Education of Parents	Dist
Input Indicators	
Expenditure per Pupil	Dist
Instructional Time	Buil
Teacher Qualification	Buil
Administrator Qualification	Dist Bui
Staff Salary Levels[3]	Dist
Staff Pupil Ratios	Buil
Curriculum Breadth	Buil
Access to Advanced Courses	Dist
Availability of Cocurricular activities	Buil
Graduation Requirements	Dist
Facility Adequacy[4]	Buil
Support Services	Buil
Auxiliary Services	Dist

[3]These items should be disaggregated to differenti
auxiliary staff.

[4]It may be useful to measure those items in terms
function.

TORS (continued)

ere ected	Where Reported	Utility	Recommen-dation
rict	Local	Evaluation	
rict	Local	Evaluation	
rict	State	Evaluation	*
ding	State & Local	Accreditation	*
ding	State	Accreditation	*
rict & lding	State	Accreditation	*
rict	State	Accreditation	*
ding	State & Local	Accreditation	*
ding	State	Accreditation	*
rict	State	Accreditation	*
ding	State	Accreditation	*
rict	State	Accreditation	*
ding	Local	Accreditation	*
ding	Local	Evaluation	
rict	Local	Evaluation	

ate teachers from specialists and support or

of percentage of budget allocated to that

Indicator	Whe Colle
Input Indicators (continued)	
Equipment Availability	Build
Supply Availability	Build
Process Indicators	
Policy Development	Distr
Planning Procedures	Distr
Curriculum Management System	Distr
Home-Community Involvement	Distr
Personnel Practices	Distr
Organization	Distr
Public Relations	Distr
Records and Reporting	Distr
Curriculum Development	Distr
Decision Making	Distr
Use of Tests	Distr
Outcome Indicators	
Student Test Data	
Norm-referenced tests	Build
Criterion-references tests	Build
SAT	Build
ACT	Build
Advanced Placement	Build
Completion Rate	Build
College Entrance	Build
Postsecondary Training	Build
College Success (First Year)	Build
Job Placement	Build

ORS (continued)

re cted	Where Reported	Utility	Recommen- dation
ing	Local	Evaluation	
ing	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	Local	Evaluation	
ict	State	Assessment	*
ing	State & Local	Assessment	*
ing	Local	Evaluation	
ing	State & Local	Evaluation	
ing	State & Local	Evaluation	
ing	Local	Evaluation	
ing	State & Local	Assessment	*
ing	State & Local	Assessment	*
ing	State & Local	Assessment	*
ing	State & Local	Assessment	*
ing	Local	Evaluation	

Indicator	Wh Coll
Outcome Indicators (continued)	
Employer Satisfaction	Buil
Community Satisfaction	Buil
Parent Satisfaction	Buil
School Climate	Buil
Discipline Incidence	Buil
% Completion of Key Subjects	Buil
Civic Involvement	Buil
Cocurricular Involvement	Dist
Community Support	Dist

ATORS (continued)

ere ected	Where Reported	Utility	Recommen-dation
ding	State & Local	Evaluation	*
ding	State & Local	Evaluation	*
ding	State & Local	Evaluation	*
ding	Local	Evaluation	
ding	State & Local	Evaluation	*
ding	State & Local	Evaluation	*
ding	Local	Evaluation	
rict	Local	Evaluation	*
rict	Local	Evaluation	

APPENDIX C

TEXTBOOK USAGE IN ELEMENTARY SCHOOLS OF EDUCATIONAL SERVICE UNIT 3

TEXTBOOK USAGE IN ELEMENTARY SCHOOLS

Textbook Company	Reading	Lanauge Arts	Mathematic
Holt, Rinehart, Winston	X*		X 4-6
D.C. Heath	X	X 7-8/1-6	X
Addison Wesley			X
Houghton-Mifflin	X	X 7-8/1-3/1-6/K-6	X
Macmillan	X	X K-2/K-6	X K-3
Scott Foresman	X K	X K-6/K-6/2-6/7-8	X
Harper & Row	X 1-6	X 4-6/3/8	
Rand McNally	X 1-6		
Ginn & Co	X 4-6/7		
Prentice-Hall	X 7-8		
Harcourt Brace	X	X K-6/3-6/7-8/K	
McGraw-Hill	X		
Scholastic Press			
SRA	X		
American Book Company			
Silver Burdette		X K-6/1-6/1-6/7-8	
Follett		X 3-6/K-6	
Allyn & Bacon			
Laidlaw		X 1-6	
Economy Press		X 1-2/1-6	
Riverside			
Entry Publishing			
Merrill		X 2-6	
McDougall Littel		X 7-8/2	

OF EDUCATIONAL SERVICE UNIT 3

s	Science	Social Studies	Health
	X 7/6/8[a]	X K-6/K-6/K-6	
	X 1-2/K-6		
	X K-6/1-3		
		X 7-8/8/1-6/5-7	X K-5/K-6
	X K/1-6		
	X 7/8		
	X K-6	X	
	X K-6		
		X K/2	
	X 8	X K/4-6	
X	1-6/4-6/K-6/7/8	X 1/K-6/1-5/1-6/1-6	X K-6
		X 7/7/K-6/6/K-6/1-6	
		X 7/K-6/6/8	
	X 4-6/8	X 8	X K-6
		X 1-4	
		X 8	
		X 7-8	
X	K-6/7/8		

37

TEXTBOOK USAGE IN ELEMENTARY SCHOOLS OF EDUCAT

Textbook Company	Reading	Language Arts	Mathematic
Modern Curriculum Press		X 1-3/1-2	
Zaner Blozer		X K-6	
Curriculum Associates		X K-6	
Sadlier		X 7-8	
Lippencott	X K-3		
SATA			

*Adopted by schools in ESU #3.

[a]Grade levels of schools which have adopted textbook

IONAL SERVICE UNIT 3 (continued)

cs	Science	Social Studies	Health
	X K-6		

APPENDIX D

PARTIAL SUMMARY OF COURSE OFFERINGS IN NEBRASKA SECONDARY SCHOOLS

PARTIAL SUMMARY OF COURSE OFFERINGS IN NEBRASKA SECONDARY SCHOOLS

The following is a summary of the frequency with which certain courses are offered in Nebraska high schools. Only a partial listing is provided, the purpose of which is to indicate the degree of dispersion.

Course Title	Percent Offered 1987-1988	Percent Offered 1988-1989
Agriculture I	39.9	40.7
Horticulture	2.0	4.3
Art I	94.7	95.0
Art III	46.8	46.7
Accounting/Bookkeeping	97.7	98.3
Business Machines	16.3	13.7
Advanced Shorthand	2.3	1.3
Beginning Keyboarding/ Typewriting	99.7	99.0
Retailing	1.0	1.3
English I	97.7	98.7
English IV	88.0	88.0
Advanced Speech	30.6	31.3
Debate	6.0	6.0
Language Arts-Social Science Core	4.0	3.0
Latin I	1.7	2.0
Spanish I	54.8	57.3
Spanish IV	9.6	10.7
Health Occupations	0.7	0.3
Physical Education I	86.7	89.0
Comprehensive Homemaking	84.4	82.3
Child Development	52.2	54.3
Food and Nutrition	69.1	64.3
Drafting	62.5	63.7
Woodworking	77.4	78.3
Graphic Arts	2.7	4.7
Algebra I	100.0	99.0
Calculus	20.3	22.0
General Music	51.8	49.3
Instrumental Music	97.3	97.0
Biology I	99.7	100.0
Physics	80.4	75.0
Social Studies	64.1	66.0
Anthropology	2.7	1.3
Economics	34.6	37.7
Automotive Services	10.0	11.0
Driver Education	38.5	36.3
Career Education	18.6	18.0

APPENDIX E

TESTING OPTIONS AND COSTS

TESTING OPTIONS AND COSTS

Introduction

Tests of all kinds are a prevalent fact of life in the American educational system. Tests are used for decisionmaking about individual students and groups of students. Decisions are made about whether or not a preschooler starts school early through individual intelligence test results providing useful information to help make that decision. Teacher-made tests are used throughout the educational system from grade school through graduate school to help instructors assess how well their students are learning on a day-to-day or week-to-week basis. Teachers use these results to adapt the pace and content of their instruction as well as to evaluate and grade student performance. National standardized tests are used for instructional decisions about who gets admitted to college, to law school, and to medical school, and sometimes they play a role in influencing who graduates. From kindergarten to graduate school, tests in one form or another affect most every student and play a significant role in helping teachers make decisions about their instruction, administrators make decisions about curriculum, and policy makers make decisions about programs and schools. Testing is more than a necessary evil in our educational system; it is a useful and needed tool.

As useful as tests can be, however, they can also be misused. Tests are not perfect and even if they were, the people who use them are not infallible. Students can have their bad test days. Distractions can affect test performance. Mistakes can be made in scoring or reporting test results. Test-givers can misread directions; test-takers can misunderstand the questions; and persons who use the tests to make decisions can misinterpret the results. When tests are used for the purposes of comparing one student to another group of students, one class to similar classes nationally, or students in one country to students in another country, the possibilities for error, confusion, and misinterpretation are dramatically increased. Too often, tests designed for one purpose are used for completely different purposes. Tests are accurate, when validated, only for the purpose they were intended. When particular tests or test results are used for another unintended purpose, the interpretation is likely to be greatly flawed. Therefore, it is important that tests are not the sole factor, especially the results of one test score, when teachers make decisions about individual students or even large groups of students. Policy makers must also be cautious when interpreting test results.

If Nebraska educational policy makers consider the possibility of a state-wide testing program, it is essential that the usefulness of a testing program be enhanced to its optimum and that misuse of testing information be minimized. To ensure that tests are used constructively to improve the education in the state and to prevent misuse of test information, key assumptions form the foundation for the thinking about a state-wide testing program expressed in this paper.

Key Assumptions

1. Tests can be a useful tool for helping improve education in the state.
2. A state-wide testing program should be a high quality program.
3. A high quality testing program demands high quality personnel and adequate resources (e.g., budget, personnel, technical assistance). A state-wide testing program must be adequately funded; it cannot be viewed as simply another add-on to expectations without supportive funding.
4. Test results should not be the sole criterion for any educational decision, whether the decision is about students, programs, or schools.
5. Tests should be one part of an evaluation system.
6. Tests need to be selected for a specific purpose and used for what they can do, not for what they cannot do.
7. Decisions about programs, curricula, or schools should not be driven solely by test results. Teachers change their instructional practice using a variety of cues as well as test results to help them determine how well students are learning. Policy makers need to do the same.
8. Test information can be misused and every effort must be made to ensure that it is not misused. Any good testing program should have safeguards against misuse whether that potential misuse involves individual students or entire school districts.

Purposes of a State-Wide Testing Program

The consideration of tests for decision making is a complicated process. While some tests can be used to provide useful information for a variety of decisions, no one single test can serve all the purposes of a state-wide program. The purposes of a state-wide program may include accreditation, state accountability,

school improvement, or student improvement. For example, if the purpose of a testing program is to provide information useful in making decisions about accreditation of schools in the state, preset criteria for acceptable performance must be established on important state-adopted objectives of education. This would necessitate the development of sets of state-wide objectives for education at various grade levels in specified content areas. A criterion-referenced test, specifically assessing performance of these prespecified outcome objectives across grade levels, would need to be developed or adopted. However, if the purpose is one of state accountability, a norm-referenced test would probably be preferable as the information needed is more one of comparability of scores from students in Nebraska schools with themselves and with those of students nationally. If a test is needed to provide information that will guide improvement decisions that are tied directly to local curriculum plans, a test that gives curriculum-based information that is directly relevant to the local curriculum is needed. Such a test would likely be developed at the local level and be a criterion-referenced instrument. Similarly, when tests are used to make individual student decisions about objective mastery or need for additional instruction, an instrument that is tied directly to the individual goals of the local school instructional program is needed.

Currently Available Standardized Tests

If the purpose of the assessment is strictly for accountability, a commercially available standardized achievement battery would possibly be appropriate. Several achievement test batteries are available that differ in content specifications, curricular objectives, and type of norm group. The choice of which test would be best for this testing purpose would necessitate a careful decision-making process that involves test specialists, school administrators, and teachers. In general, these instruments have been shown to possess adequate psychometric qualities.

If the purpose of testing is such that a criterion-referenced approach is appropriate, there are several commercially available tests that might serve this purpose. One often-mentioned possibility is the state testing program option offered by the Natoinal Assessment of Educational Progress (NAEP). This testing program has an item pool which has been administered in the national test program; therefore, some historical item performance information from a national representative sample is available.

Based on a recent communication with an associate of the NAEP program at the Educational Testing Service, the availability of the NAEP program for state-wide customized test development

is very limited and under transition. NAEP is currently involved in a national effort to develop a national performance-based outcomes assessment system that is designed to allow for comparisons across individual states. Because of this new endeavor, NAEP has dramatically reduced the available item pool for individual state testing programs. At this time, this does not seem to be a feasible option for state performance-based test development. Even more important is the lack of information on the match of test item objectives to the state performance-based assessment objective for the State of Nebraska.

There are several commercially available standardized criterion-referenced tests. A listing of these tests, the content areas, grade levels, and price is contained in an addendum to this report. In addition, a summary of a qualitative review from the *Mental Measurements Yearbook* for each test is presented. In general, these tests have been rated to be poor in psychometric quality and not generally relevant to specific outcome-based performance objectives. Therefore, these tests do not seem appropriate for use in a state-wide performance-based testing program.

Implementation of a State-Wide Testing Program

Assuming that a decision is made to implement a state-wide testing program, a description of such a program and its associated costs is impossible until the purposes of testing become more clear. A testing program cannot meet all of the goals described earlier without involving multiple tests with considerable cost to the state. In addition, there are several critical issues surrounding implementation of a testing program.

Selection of the Test

A major decision to be made is whether to develop one's own test or to use a currently existing standardized test. Essentially, this decision becomes one of degree to which an existing test matches chosen curriculum goals. If there is a sufficient match, the use of an existing test is most economical. It is, however, quite possible that a "home-grown" test will need to be developed; if existing tests do not match the state's curriculum goals, the state should devleop its own test.

If should be realized that a locally developed Nebraska test would involve a great deal of resources, particularly during the startup phase of the program. Moreover, it appears that the State Department of Education would need additional personnel with testing expertise to begin such a program. Hence, in order

to effectively initiate a testing program, the state would have to either hire several new staff members with strong test development skills or contract out the test development to qualified individuals or organizations.

Matching the Test to the Curriculum

Central to the issue of test-curriculum fit is the identification of the common curriculum in Nebraska schools. The idea that a fair test would be one that measures only those objectives that are common to all schools sounds deceptively simple. It is an idea that has serious and complicated implications for the future of education in Nebraska. A major concern is that the common core of the curriculum consists of general basic skills (e.g., reading, math) and that the resulting test will measure only those skills. There is a fear that such a test would drive school curricula away from the thinking of things that are not part of the common core. Indeed, the Policy Study Task Force has found clear indications that there is a tremendous amount of diversity in the curricula across the state. This suggests that what is "common" may very well be only the most basic academic skills.

Implementing a common test across the state would have an important political implication--the local control of districts to choose their own curricula would be seriously threatened. There is a major concern that if a single standardized test is used state-wide and the test scores are used to compare districts, then a "homogenization" of the curricula across districts would result. That is, districts would be motivated to alter their curricula to best suit the test upon which the comparisons would be made. This is a very real problem; there is evidence that such a homogenization of curricula has occurred in a number of other states.

Another testing program option would be to have districts test over unique aspects of their curricula as well as the common core curriculum across the state. This type of test customization would help allow districts to maintain their unique curricula, and thus encourage continued curricular diversity across districts.

Who Will Be Tested?

The issue of who will be tested is basic to consideration of a state-wide testing program. Will every student be tested or some sample from each school or district? The answer again depends upon the purpose of the testing program. If the purpose is accreditation or accountability, a random sample of students from each district should be satisfactory. However,

if the purpose is for school or student improvement, each student should probably be tested.

Testing Program Costs

Given the uncertainty regarding the issues discussed above, it should be clear that it is impossible to provide an estimate of the costs of a state-wide testing program without several assumptions being made. In this section, two hypothetical programs will be described along with their corresponding cost estimates. Note that these are examples designed to provide a perspective on program costs, and are not testing programs that we are necessarily recommending.

Suppose that the decision had been made to develop a testing program in which three content areas would be measured in three separate grades. Further, assume that every Nebraska student in these grades is to be tested, meaning that roughly 25,000 students per grade would be tested.

If the decision was made to develop the test locally, there would need to be item pools of at least 200 items for each of the nine tests (3 areas x 3 grades). The estimated startup costs of such a program would be approximately $500,000 for the first year. Moreover, in each subsequent year there would be costs associated with printing, distributing, scoring, and reporting of tests results, as well as the development of new test items. Maintaining the test program would cost approximately $250,000 per year.

If the decision were made to use an existing standardized test, there would be substantial savings in the startup costs, since item pools would not have to be developed. There would be, however, the ongoing costs of printing (or purchasing) test forms, test distribution, scoring, and reporting of results. Again, the costs would be in the neighborhood of $250,000 per year.

These cost estimates are based upon our assessment of what it has cost other states to initiate and maintain state-wide testing programs. Note that if the assumptions are changed (e.g., testing a sample of students rather than all students) the cost estimates would change. What must be stressed, however, is that the implementation of a credible testing program is not an inexpensive endeavor. There must be a clear commitment of funds to such a project in order for it to provide useful information. In essence, a poorly developed or inadequately funded testing program would probably be worse than no program at all.

Considerations

On the basis of the noted information, we make the following considerations:

1. A decision needs to be made about the specific purpose and expected uses of the testing program. The purpose should be stated with as much clarity and precision as possible. (As noted, all subsequent decisions emanate from this decision; see Recommendations 4 and 5, p. 19).

2. Implementation issues, which will flow naturally from decisions made about the purposes of a testing program, need to be discussed and procedures developed for conducting and monitoring the program. These implementation issues include such matters as determining included or excluded (e.g., special education) students, and how test results will be recorded and reported. (See Recommendations 3 and 11, pp. 18 and 21.)

3. Technical and implementation decisions about test selection and sampling procedures, for example, should be made using available consultation resources in the state (e.g., Buros Institute of Mental Measurements, which is a national center on tests).

4. A commitment must be made to the development of a high quality testing program, regardless of what kind of tests are used or what the purpose of the testing program is.

5. Funding levels need to be appropriate to support a high quality testing program with high quality staff. (We do not believe that the needed expertise to develop and maintain a state-wide testing program is currently available; see Recommendation 3, p. 18).

6. A testing program needs to be considered as only one part of a larger evaluation system which includes numerous evaluative indicators that provide information for policy makers who make decisions and recommendations about school improvement and improvement of education in Nebraska (see p. 5 of text).

NATIONAL ASSESSMENT OF EDUCATION PROGRAM (NAEP)

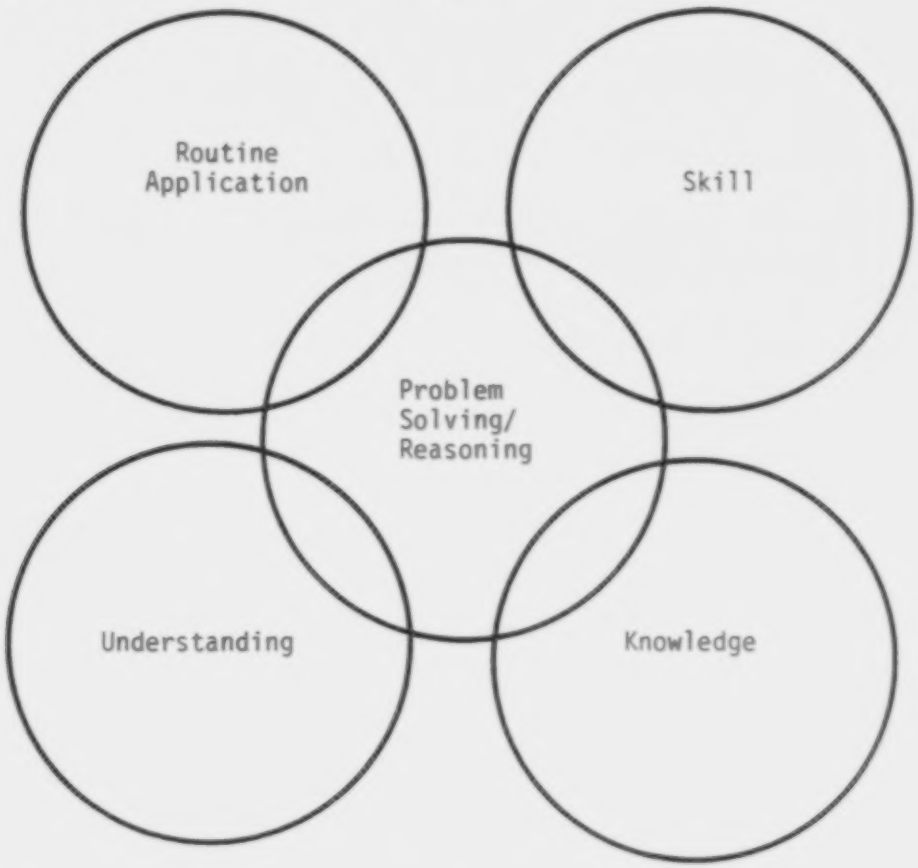
Math Objectives: 1985-1986 Assessment

Introduction and Overview

In 1969 the National Assessment of Education Progress (NAEP) was established for the purpose of collecting information about the performance of 9-, 13-, and 17-year-old Americans in a variety of subject areas.

Assessments in mathematics were conducted during the school terms ending in 1973, 1978, and 1982, and a fourth is being held during 1986.

There are five mathematics objectives:



Since all of the objectives involve elements of problem solving and since the Assessment Policy Committee requested an emphasis on critical thinking and problem solving, the relationship among the five process areas of the objectives can be seen in the above diagram.

Categories of Mathematics Content

The categories consist primarily of elementary and secondary school mathematics up to but not including calculus.

1. Fundamental Methods of Mathematics:

   Models, Problem-Solving Strategies, Informal Induction, Informal Deduction, Routine Procedures, Axiomatic Systems Logic

2. Discrete Mathematics:

   Probability, Permutations and Combinations, Matrices

3. Data Organization and Interpretation:

   Organizing, Displaying, and Interpreting Information, Measure of Central Tendency, Measure of Spread and Position Sampling and Polling

4. Measurement:

   Concepts of Measurement, Standard Unit of Measure, Measuring Instruments, Applications of Measurement

5. Geometry:

   Geometric Figures and Their Properties, Relations Established by Definition, Relationships Established by Theorem, Motion Geometry, Coordinate Systems Applied to Geometry, Spatial Visualization

6. Relations, Functions, and Algebraic Expressions:

   Use of Variables, Simple Relations and Functions, Solving Equations and Inequalities, Coordinating Systems, Functions, Trigonometry

7. Numbers and Operations:

   Numeration, Number Concepts, Operations, Mental Computation, Estimation, Properties, Relationships

Not all of the subtopics are tested at every age level.

Brief Description of Skills and Abilities Needed in Problem Solving

1. Pre-analysis Skills

2. Analysis Skills

3. Ability to Select Viable Strategies for the Solution of a Given Problem

4. Ability to Interpret the Solution of a Problem

Process	Age 9	Age 13	Age 17
a. Problem Solving/Reasoning	17	18	19
b. Routine Applications	12	15	15
c. Understanding/Comprehension	19	25	28
d. Skill	40	32	29
e. Knowledge	12	11	10

*Percents may not total 100% due to rounding

Figure 1

Approximate Percentage Distribution of Exercises by Age and Process

Content	Age 9	Age 13	Age 17
a. Mathematical Methods	12	9	10
b. Discrete Mathematics	2	4	5
c. Data Organization/Interpretation	11	8	6
d. Measurement	19	17	13
e. Geometry	5	9	11
f. Relations, Functions, etc.	5	5	13
g. Numbers and Operations	47	49	43

*Percents may not total 100% due to rounding

Figure 2

Approximate Percentage Distribution of Exercises by Age and Content

Calculator Usage

The use of the calculator has been included in NAEP beginning with the second assessment.

The five categories are:

1. Routine Computations
2. More Difficult Computations
3. Understanding Concepts
4. Exploration
5. Applications and Problem Solving

Exercises in some categories, such as understanding concepts and exploration, are more appropriate for instructional purposes than for assessment. Therefore categories 1, 2, and 5 are measured in the greatest depth.

Some exercises are given with and without calculators to provide data on student performance with and without the aid of the calculator.

Summary of Assessment Procedures:

Each of the four mathematics assessments (1973, 1978, 1982, 1986) involved nationally representative samples of 9-, 13-, and 17-year-olds, and together the assessments generated data from a total of 150,000 students for the examination of trends.

In the 1986 assessment, NAEP sampled students by grade as well as age, making available a second data set based on 34,000 additional students in grades 3, 7, and 11.

The mathematics assessments include both open-ended and multiple-choice questions covering a wide range of content and process areas.

Based on the assessment results, five levels of proficiency have been defined:

Level 150 - Simple Arithmetic Facts
Level 200 - Beginning Skills and Understanding
Level 250 - Basic Operations and Beginning Problem Solving
Level 300 - Moderately Complex Procedures and Reasoning
Level 350 - Multiple-step Problem Solving and Algebra

NAEP's mathematics scale was computed as a weighted composite of proficiency on five content area subscales. Thus, results are also available indicating students' relative strengths and weaknesses across these content areas.

Reading Objectives: 1986 and 1988 Assessments

Introduction and Overview

The central attention accorded to reading reflects the fact that, although it has no substantive content of its own, reading proficiency as a process plays an important role in every subject-matter area.

As a whole, the set of objectives is based on an interactive view of reading.

The charge given to NAEP is that assessment objectives should reflect the thinking of a wide variety of individuals, state and district curriculum specialists, teachers, school administrators, researchers, parents, and concerned citizens, officials and business people.

While objectives defined from such a consensual process cannot specifically advance either a single theoretical framework of views of any one individual, they do represent the thinking of a broad cross-section of individuals who are deeply concerned with reading in our schools.

Objectives:

There are four objectives as follows:

1. Comprehends what is read:

    a. Comprehends various types of written materials
    b. Comprehends materials read for a particular purpose

2. Extends comprehension:

    a. Analyzes what has been read
    b. Inteprets what has been read

3. Manages the reading experience:

    a. Uses the structure and organization of the text
    b. Uses reader's aids
    c. Shows flexibility in approach to reading
    d. Selects reading materials appropriate to the purpose

4. Values reading:

    a. Values reading as a source of enjoyment
    b. Values reading to expand understanding and fulfill personal goals

c. Values reading as a means of acquiring knowledge and learning new skills
d. Values the cultural role of written language

Instructional Strategies

There are five instructional strategies to help teachers who choose to use or adapt the NAEP reading objectives:

1. Reading Experiences: Answering/Posing Questions, Integrating Information, Clarifying/Speculating
2. Linking Reading to Writing
3. Relating to Information/Reading One Work to Another
4. Helping Students Find Information
5. Helping Students Understand Text Structures

NAEP's Reading Proficiency Scale

1. Background:

   In each of the NAEP's assessments of reading, students have been asked to respond to multiple-choice questions, to answer brief, open-ended questions, and to write about their reactions to what they read.

   The common goal of each assessment has been to understand and present information about the status and progress of reading proficiencies among selected school-age populations.

2. Description:

   NAEP has used the item response theory to enhance the comparability of results across ages, groups, and time.

   The NAEP reading scale reflects proficiency in comprehending or constructing meaning from a broad range of prose materials.

   An innovative process of anchoring the scale was undertaken to enhance it interpretability. This anchoring process was aimed primarily at guiding the interpretation of performance at specific scale points, and thus toward establishing the construct validity of the scale.

   NAEP identified five scale points at 50-point intervals, selected tasks that discriminated well at these repsective levels, and analyzed performance both in terms of the test and the accompanying exercise.

Levels of Proficiency (5)

1. Rudimentary (150)
2. Basic (200)
3. Intermediate (250)
4. Adept (300)
5. Advanced (350)

Science Objectives: 1985-1986 Assessment

Objectives Development Process

NAEP science objectives have been developed as a framework for measuring the attainments in science of students in the United States at ages 9, 13, and 17.

The objectives (resulting from a consensus process involving many people) for 1985-1986 build on what has gone before, taking into account recent developments and new emphasis.

Framework for Science Objectives

A matrix in three dimensions--Content, Context, and Cognition--represents the broad objectives of science education.

A national assessment of science can cover only a small sample of the domain included in science education throughout the school years. Therefore, the committee needed to consider how much weight to give each of the major categories in the matrix.

Content	Age 9	Age 13	Age 17
Life Sciences	35%	25%	25%
Physics	15%	15%	20%
Chemistry	10%	15%	20%
Earth and Space Science	20%	20%	15%
History of Science	5%	10%	5%
Nature of Science	15%	5%	15%
	100%	100%	100%

Context	Age 9	Age 13	Age 17
Scientific Personal Societal Technological	50%	55%	60%
	50%	45%	40%
	100%	100%	100%

Cognition	Age 9	Age 13	Age 17
Knows	40%	40%	40%
Uses	30%	30%	30%
Integrates	30%	30%	30%
	100%	100%	100%

Attitudes, Values, and Experience

The collection and reporting of information about students' attitudes and values with respect to science and scientific endeavors are important parts of the assessmnet.

The assessment monitors trends by repeating some questions from previous assessments and reporting how students' attitudes have changed over the years.

Seven categories of questions for self-reporting by students are:

1. Attitudes Toward Science Classes
2. Career and Education Intentions
3. Socioscientific Responsibility
4. Science as a Personal Tool
5. Value of Science
6. Societal Issues
7. Experience in Science

U.S. History Objectives: 1988 Assessment

Introduction

The assessment of students' understanding and knowledge about United States history represents a new subject area for NAEP (at the three age/grade levels 9-, 13-, and 17-year-olds, as well as fourth, eighth, and twelfth grades)

The assessment was developed through a comprehensive review process.

Objectives Framework

The framework for what young people should understand about the American past is organized in three dimensions, each of which constitutes an objective for the assessment:

1. Chronology of events, persons, and documents

2. Context of American political, economic, cultural, social, family, and intellectual life within each chronological period and how these interact and change over time

3. Reasoning Skills necessary to understand how the record of the past is interpreted and reconstructed

Civics Objectives: 1988 Assessment

Introduction

Civics was first measured in the context of the 1969-1970 citizenship assessment, again in the 1974-1975 citizenship assessment, and a third time in 1982-1983 as part of a combined assessment of citizenship and social studies.

The objectives represent a modification of some of the 1981-1982 citizenship objectives and additional objectives which reflect current trends in civics education.

Objectives Framework

A three-dimensional matrix made up of Content, Context, and Cognition outlines the broad objectives for civics education.

These dimensions help define the major objectives of United States civics education and provide guidelines for developing questions to assess students at age 9/grade 4, age 13/grade 8, and age 17/grade 12.

The three objectives are:

1. Context: Home, School, Community, State, Nation, World

2. Cognition: Knows, Understands, and Applies

3. Content: Right, Responsibilities, and the Law; Political Processes; Political Institutions (Structure of Foundations); and Democratic Principles and Purpose of Government

Geography Objectives: 1988 Assessment

Introduction

To measure and report on the extent and quality of students' understanding of geography, the National Geographic Society initiated and is providing support for a national assessment of students 17 years old and in the twelfth grade.

The assessment is the first in NAEP's 20-year history.

The objectives are a consensus of opinions and priorities that emerged from the deliberations of those involved.

The objectives are not intended as complete or definitive specifications of curriculum topics. Rather, they provide an overview of learning outcomes.

The final selection of exercises for the assessment involved careful consideration of the following requirements:

1. The need to provide exercises across a wide range of difficulty
2. The need to vary the question format
3. The need to meet the distribution percentage specifications designated by the Learning Area Committee

Framework for the Geography Objectives

The framework for what twelfth-grade students and 17-year-olds should understand in geography has the following dimensions:

1. Geographic Skills and Tools: Using Maps, Using Charts, Using Globes
2. Geographic Knowledge and Concepts (Content): Understanding Physical Geography, Understanding Cultural Geography
3. Geographic Inquiry: Knowing, Understanding, and Applying

A Framework for Assessing Computer Competence: Defining Objectives

Background

The 1986 computer competence objectives are merely a guide for designing the first national assessment of computer competence.

The development process relied upon subject area specialists to ensure that the objectives reflect both current practice and expert opinion.

The objectives are the result of a consensus process incorporating the opinions of concerned educators and citizens representing various backgrounds and points of view.

This assessment will measure students' abilities in areas that are not currently in the curriculum of many schools and districts. This approach is consistent with NAEP's purpose: to find out what students know and to track changes in their knowledge over time, rather than to measure districts, schools, or students against some norms.

Overall Plan for the Assessment

1. Defining computer competence:

   It is important to report on students' computing skills in a way that embraces both the view that students should be learning the rudiments of computer science and the view that students should be able to use computers for a variety of purposes.

   The purpose of the assessment is not to assess the value to schools of computers as instructional delivery systems, but rather the competence students exhibit as users of computers.

2. Measuring computer competence:

   Central issue: How to determine whether someone is a competent user of computers

   The variety of computers, operating sytems, and peripheral equipment in use, and the differences among them, are simply too great to permit the writing of questions and scoring of responses that could determine whether students know how to set up and operate a system.

   In these situations, it is preferable to use performance testing to get a the answers to such questions, using whatever equipment is normally available to students.

Two Issues:

a. Given an exercise in which a paper-and-pencil test might plausibly be a valid substitute for a performance test, is it in fact a valid substitute?
b. In some cases no plausable paper-and-pencil test for a given competence can be imagined, let alone validated.

However desirable performance testing might be in the light of the limitations of paper-and-pencil tests, issues of validity and reliability do not disappear when performance tests are used, because performance contexts may differ widely and observers may make very different judgments on the same performance.

3. Three Major Categories of Objectives:

The 1986 computer competence assessment will be composed of 309 different exercises to be administered across three age and grade levels of students (age 9/grade 3; age 13/grade 7, age 17/grade 11).

Computer Objective Categories:

1. Computer Application. Will be distributed among the cognitive levels such as knowledge, operation, problem solving/design (heavily assessed at age 17/grade 11 level)

    a. Word Processing
    b. Database Management
    c. Lab Instrumentation
    d. Telecommunications
    e. Graphics
    f. Music Generation
    g. Spreadsheets
    h. Models and Simulations

2. Computer Science Programming. Will be distributed among the cognitive levels such as knowledge, operation, problem solving/design (heavily assessed at age 13/grade 7)

    a. Elements of Programming Language
    b. Structures of Data and Control
    c. Program Planning Design and Testing

3. Knowledge and Attitudes

    a. Systems
    b. History
    c. Uses
    d. Beliefs

The number of questions in each of the major categories will not be the same at each age/grade level.

There will be extensive overlap across levels. This will enable analysts to compare the performance of students at different age/grade levels on identical exercises.

ACHIEVEMENT BATTERIES

Assessment of Basic Competencies (ABC)

Ages: 3+ to 15
Publication Year: 1981
3 Subtests (Information Processing, Language, Mathematics); 2 response forms (Diagnostic Edition, 12 pages; Developmental Edition, 12 pages)
Price Data: (1983)
$414 per starter set including all materials necessary for administering all 3 subtests; $13.50 per 20 response forms of either type; $24 per manual; prices for training package available upon request.
Publisher: Scholastic Testing Service, Inc.

Reviewer 1:

a. Reliability--for grades K, 2 and 6 = .76 to .98
b. Validity: Needs to be demonstrated in all areas and for all groups. The manual suggests it is appropriate to make inference.
c. Recommendation: ABC cannot be recommended strongly for use.

Reviewer 2:

a. Reliability--for grades K, 2 and 6 and K to 8 combined:
(Broad domains = .95 to .96)
(Subtests = .88 to .95)
b. Validity: Documentation of content validity is weak. Criterion-related validity is generally satisfactory.
c. Recommendation: Limited evidence of validity and the lack of a clear relationship between item content and curricular material. Cannot be recommended without reservation.

Brigance Diagnostic Inventory of Basic Skills (IBS)

Grades: K-6
Publication Year: 1976
141 Specific-objective tests in 4 areas: Readiness (24 tests) Reading (33 tests), Mathematics (64 tests), Language Arts (20 tests)
Price Data: (1985)
$69.95 per examiner's kit including manual, test, and 10 student record booklets; $11.95 per package of 10 student record booklets.
Publisher: Curriculum Associates, Inc.

Comments: No suggested standards of mastery for 65% of the tests.
No specific time limits are listed on many tests; others are untimed.

Reviewer 1:

1. Well-suited to determining mastery of very specific teaching objectives
2. Offers the teacher an excellent model for pretesting and posttesting for mastery of actual classroom curriculum objectives
3. Due to a normative or research sample, it is not appropriate for making decisions that compare children against one another on specific academic materials
4. The measure is best used in conjunction with evaluations of the student's classroom products, classroom observation, and scrutiny of actual curricular goals

Reviewer 2:

1. The IBS does have a number of commendable qualities which include good content validity, ease and flexibility in administration, and an excellent format for reporting results and record keeping
2. The most critical problem is a lack of attention to practically every important technical detail in test development
3. The greatest deficiency of the IBS is the complete absence of data concerning reliability and validity

Brigance Diagnostic Comprehensive Inventory of Basic Skills (CIBS)

Grades: K-9
Publication Year: 1983
198 Specific-objective tests in 7 areas: Readiness, Reading, Listening, Research and Study Skills, Spelling, Lanuage, Math
Price Data: (1983)
$99 per inventory and 10 record books; $15 per 10 record books; $6.95 per comprehensive class record book; $15.95 per 30 objective forms
Publisher: Curriculum Associates, Inc.
Comments: No standards of mastery for 20% of the tests
Alternative forms (A and B) for 51 of the tests
Specific time limits are limited on many tests, others are untimed

Reviewer 1:

1. Content descriptions are not available with the Inventory, and no statements are made indicating that they were used in test development
2. Specific placement of skills in various popular texts is not referenced
3. No data on item quality are provided
4. No reliability and validity data are reported
5. Application of cutoff scores is not clear in all areas
6. Lack of information--limits the generalizability of results and the usefulness of the tests

Reviewer 2:

1. Well-constructed inventory that is easy to administer and score and covers a wide age range
2. It does not tell the test user how good or poor a subject's achievement is in each of the skill areas
3. CIBS should therefore be used in conjunction with good norm-referenced tests

Brigance Diagnostic Inventory of Essential Skills (IES)

Grades: 4-12
Publication Year: 1981
191 tests in 4 areas: Reading (30 tests), Language Arts (25 tests), Mathematics (69 tests), Life Skills (67 tests)
Forms A and B for 6 of the tests, all in the same books
Price Data: (1983)
$99.95 for examiner's kit including manual, tests, and 10 student record books; $16.95 per package of 10 student record books
Publisher: CUrriculum Associates, Inc.
Comments: Specific time limits are listed on mnay tests; others are untimed

Reviewer 1

1. Very detailed criterion-referenced test aimed at secondary school students and designed to assess basic skills that are seen as necessary for successful functioning as an adult
2. Test can form a useful part of an Individual Educational Plan for a child whose high school education is focused on acquiring basic skills
3. Author of the IES clearly recognizes the limitations of criterion-referenced tests, and the manual is explicit in warning against some possible misuses of the results

4. IES cannot be used as part of a single assessment session; the overall administration time for the entire instrument would be many hours
5. Its usefulness is more within a classroom framework than within the more typical diagnostic framework of an individual assessment
6. For classification it has little utility since neither norms nor reliability data are provided
7. Its use is also primarily with children for whom minimum survival skills are the educational goal, not for high school students functioning at an average level or above

Reviewer 2:

1. The use of this instrument should be precluded in special education settings until its psychometric properties have been empirically demonstrated

Canadian Achievement Tests (CAT, Form A)

Grades: 1.6-2.9; 2.6-3.9; 3.6-4.9; 4.6-5.9; 5.6-6.9; 6.6-7.9; 7.6-9.9; 9.6-12.9
Publication Years: 1981, 1983
8 Overlapping Levels:
- a. Locator Tests--Grades 2-6, 6-12
- b. Level 12--Grades 1.6-2.9
- c. Level 13--Grades 2.6-3.9
  9 Subtests for Levels 12 and 13: Reading (4 tests), Spelling (1 test), Language (2 tests), Math (2 tests)
- d. Level 14--Grades 3.6-4.9
- e. Level 15--Grades 4.6-5.9
- f. Level 16--Grades 5.6-6.9
- g. Level 17--Grades 6.6-7.9
- h. Level 18--Grades 7.6-9.9
- i. Level 19--Grades 9.6-12.9
  8 Subtests for Levels 14, 15, 16, 17, 18, and 19; Reading (2 tests), Spelling (1 test), Language (2 tests), Math (2 tests), Reference Skills (1 test)

Price Data: (1984)
$42.85 for 35 reusable test booklets of any one level (grades 3.6-12.9); $10 per test coordinator's handbook; $12.50 per technical bulletin
Publisher: Canadian Test Centre/McGraw-Hill Ryerson Ltd.
Comments: Subtests available as separates

Reviewer 1:

1. Constitutes a unique approach to educational evaluation, since they merge two approaches to achievement measurement: norm-referenced and criterion-referenced assessment
2. Provincial and school district curriculum guides, major textbooks, and the objectives of several criterion-referenced testing programs were used as the basis for writing test or category objectives
3. Test Content:
   a. The number of items in each area assessed is frequently insufficient to provide detailed data necessary for individualized education programming
   b. If the publishers' claims that the tests can serve as a criterion-referenced test are to be justified, this limitation should be corrected and data provided to validate the usefulness of the test for individualized programming
   c. These can be overcome by using subtests from other batteries (e.g., in assessment of reading skills), when necessary for individual children

The Child Center Operational Assessment Tool (OAT)

Grades: Regular or special classroom students
Publication Years: 1971, 1977
4 tests: Reading/Spelling test, Math test (2 forms), Language test, Behavioral questionnaire
Price Data:
$.95 per spelling manual; $3.50 per instruction book; $.95 per math form; $.95 per test; $.95 per questionnaire
Publisher: The Child Center
Comments: Criterion-referenced; assesses and identifies children with special educational needs; administration time not reported

Reviewer 1:

1. No intended age or grade level
2. No suggestion for the approximate curriculum levels of the content sampled by the various tests
3. No specific reference
4. No concrete data of the basis for establishing the criterion scores
5. Difficult to argue that these tests are really "tests" at all

Reviewer 2:

1. Poorly manufactured
2. Lacks any information related to norm group characteristics, reliability, or validity data

Criterion Test of Basic Skills

Grades: K-8
Publication Year: 1976
2 Tests: Reading (19 specific objective subtests in 6 areas), Arithmetic (26 specific objective subtests in 12 areas)
Price Data: (1984)
$33.50 per set of testing materials including manual, stimulus cards, 25 record sheets for each test, and 25 math problem sheets in a vinyl folder
Publisher: Academic Therapy Publications
Comments: Mastery level defined as 90-100% correct; Instructional level defined as 50-80% correct; Frustration level defined as 0-49% correct
No data on reliability
No norms
No information presented in support of increasing difficulty levels (student takes only subtests beginning with 2 mastery scores and ending with 2 frustration scores) or relevancy to grade levels
Reviews: None available

Curriculum-Referenced Tests of Mastery (CRTM)

Grades: 1, 2, 3, 4, 5, 6, 7, 8, 9-10, 11-12
Publication Years: 1983, 1984
10 Levels: Level E (Grade 1) to Level R (grades 11-12)
Price Data: (1984)
$40.25 per 35 machine-scorable tests plus manual for Levels E, F, and G; $33.25 per 35 reusable tests plus manual for Levels G-R; $4.50 per manual; $5.95 per technical report; $5.50 per custom manual
Publisher: Charles E. Merrill Publishing Company
Comments: Customized tests available in which the user can tailor the test objectives to meet the curriculum
Criterion-referenced scores with suggested standards of mastery
Norm-referenced scores for reading/language arts and mathematics

Reviewer 1:

1. In its current form, the CRTM techical data are incomplete at this time
2. In terms of Reading and Mathematics achievement, the CRTM does certainly yield more specific domain- (criterion) referenced information than its competitors
3. Does not measure science and social studies (until ninth-grade level)
4. The custom edition of the CRTM is an attractive alternative for developing a local reading and mathematics test that also have national norms

Reviewer 2:

1. CRTM attempts to be both a criterion-referenced and norm-referenced test
2. If a criterion-referenced measure is needed, CRTM is an excellent candidate, especially if a custom edition is desired
3. The number of options per item ranges from 2 to 4, and this is a disadvantage in competing for reliability with other norm-referenced batteries using 5 options

Diagnostic Skills Battery (DSB)

Grades: 1-2, 3-4, 5-6, 7-8
Publication Years: 1976, 1980
4 Levels: Level 12 (Grades 1-2) 2 scores: Reading, Mathematics (Forms A, B)
Level 34 (Grades 3-4) 3 scores: Reading, Language Arts, Mathematics (Forms A, B)
Level 56 (Grades 5-6) 3 scores: Reading, Language Arts, Mathematics (Forms A, B)
Level 78 (Grades 7-8) 3 scores: Reading, Language Arts, Mathematics (Forms A, B)
Price Data: (1983)
Level 12--$2.40 per pupil for lease/store program, Plan I; $2.01 per pupil for lease/store program, Plan II
Each program has 2 scoring service plans
Level 34--$1.09 per pupil for lease/store program, Plan I; $1.60 per pupil for lease/store program, Plan II; $.95 per pupil for school purchase; $26 per 20 test booklets and one manual; $10 per 50 answer sheets
Level 56--$1.09 per pupil for lease/store program, Plan I; $1.60 per pupil for lease/store program, Plan II; $.95 per pupil for school purchase; $26 per 20 test booklets and one manual; $10 per 50 answer sheets

Level 78--$1.09 per pupil for lease/store program, Plan I; $1.60 per pupil for lease/store program, Plan II; $.95 per pupil for school purchase; $26 per 20 test booklets and one manual; $10 per 50 answer sheets
Publisher: Scholastic Testing Service, Inc.
Comments: Criterion- and norm-referenced test

Reviewer 1

1. Most school systems today are in need of both norm-referenced and criterion-referenced information for use in planning education programs
2. The DBS is designed to meet these needs
3. Reliability of the DBS: Battery composites (.93 to .97); individual test total scores (.83 to .94); part scores (.54 to .90)
4. Validity of the DBS: Battery composites (.65 to .97); reading (.55 to .97); mathematics (.58 to .96); language arts (.54 to .98)
5. It seems doubtful that the DBS could provide objective-referenced information of much use. This is not only a direct criticism of the DBS, but a concern with making inferences about mastery of skills based on a restricted number of items. Furthermore, not only is there a limited number of items but a limited number of objectives as well
5. It seems logical that the best way to conduct criterion-referenced testing would be for the school district to identify its objectives in each skill area and then develop tests of those objectives based on the curriculum and materials used in that district
6. The DBS would almost inevitably fall short of meeting the particular needs of any school district with interest in objective-referenced assessment

Distar Mastery Tests (DMT)

Grades: Preschool-Grade 3
Publication Years: 1977, 1978
6 Tests: Reading (I, II); Languages (I, II); Arithmetic Problems (I, II)
Price Data: (1984)
$35.65 per kit including 25 tests, 25 practice sheets, and examiner's manual for any one test
Publisher: Science Research Associates, Inc.
Comments: Criterion-referenced test; determines mastery for the Distar Reading (I, II), Languages (I, II), and Arithmetic Problems (I, II); no data on reliability and validity; no norms; 30-60 minutes per test

Reviewer 1

1. Because the tests are intended for users of the Distar materials to learn which of the program objectives a student has mastered, and not to compare a student's performance with others, it is reasonable that no norms are provided
2. Item development not explained in examiner's manual
3. Not clear whether the tests are designed as end-of-course, end-of-grading period, or end-of-unit tests
4. No data regarding item or test validity
5. No information on alternative form reliability
6. No discussion to explain the manner in which cutoff scores were established
7. Directions for preparing documents for scoring appear only in the manuals of Language I, Reading I, and Arthimetic II
8. Format of manuals inconsistent
9. Guidelines for interpreting test results at either the individual student or class level are not provided

IOX Basic Skill System

Grades: End of Grades 5 or 6, Grades 9-12
Publication Years: 1978, 1979
2 Levels: Elementary (End of Grades 5 or 6); 3 tests: Reading Writing, Mathematics (Forms A, B)
Secondary (Grades 9-12); 3 tests: Reading, Writing, Mathematics (Forms A, B)
Price Data:
$37.50 per 25 tests for Elementary Level (specify Form A, Form B); $42.50 per 25 reading tests for Secondary Level (specify Form A, Form B); $37.50 per 25 writing or mathematics tests for Secondary Level (specify Form A, Form B)
Publisher: IOX Assessment Associates
Comments: "Minimal competency assessment"; criterion-referenced test; 40-45 minutes per test

Reviewer 1:

1. Reliability: Internal consistency reliability is relatively low. Test-retest reliabilities = Reading (.40-.82), Writing (.63-.81); Mathematics (.65-.87)
2. Validity: Content validity and curricular relevance (domain-selection) validity are provided in manual

No evidence of construct validity
All decisions about cutoff scores and further testing are left to the user of the IOX measure

SRA Survival Skills in Reading and Mathematics

Grades: 7-12
Publication Year: 1976
2 tests: Mathematics (20 objectives, 3 items per objective) and Reading (20 objectives, 3 items per objective)
Price Data: (1984)
$27.45 per 25 tests and examiner's manual
Publisher: Science Research Associates, Inc.
Comments: Objective-based tests in reading and mathematics relating to skills essential for everyday living; no data on reliability or validity; no norms; 60 minutes per each test

Reviewer 1:

1. Tests are designed to be criterion-referenced, although little is said about how criteria were established
2. The "real world" nature of items, purportedly measuring mastery of basic skills that are critical for survival in today's world
3. Reliability procedures that have relevance for criterion-referenced tests are also absent
4. No validity evidence
5. Construct of survival is not developed as a foundation for test construction
6. If needed documentation congruent with sound measurement practices is developed, the test could provide the basis for a useful objective-based assessment of survival skills
7. In current state, not recommended

Cognitive Skills Assessment Battery, Second Edition (CSAB)

Grades: PreK-K
Pubilcation Years: 1974, 1981
98 item scores (49 consist of plus or minus) in 18 areas: Basic Information (4 scores), Identification of Body Parts (4 scores), Color Identification (4 scores), Shape Identification (4 scores), Symbol Discrimination (10 scores), Visual-Auditory Discrimination (6 scores), Auditory Discrimination (6 scores), Number Knowledge (10 scores), Letter Naming (2 scores), Vocabulary (6 scores), Information from Pictures (4 scores), Picture Comprehension (4 scores),

Story Comprehension (4 scores), Multiple Directions (4 scores), Large Muscle Coordination (4 scores), Visual-Motor Coordination (6 scores), Memory (8 scores), Response during Assessment (8 scores)
Price Data: (1983)
$48.95 per testing materials including 30 response sheets
Publisher: Teachers College Press
Comments: Criterion-referenced test; individual test; 1 form; 20-35 minutes

Reviewer 1

1. Surveys children's performance in 5 areas: Orientation Toward One's Environment, Discrimination of Similarities and Differences, Comperehension and Concept Formation, Coordination, and Immediate and Delayed Memory
2. Criterion levels are not set, and total scores are not obtained
3. Test content appears very similar to teaching materials that are presented to preschoolers and kindergartners
4. Exact item selection criteria are not given in manual (definite shortcoming) since item content is a major issue for criterion-referenced tests
5. The CSAB normative sample is not representative since only 16 communities participated in the norming study
6. Reliability data are provided from a single study of score stability over 2 to 3 weeks for 48 children
7. Scorer reliability is particularly crucial since relatively unsophisticated examiners administer the CSAB
8. Traditional validation studies are solely lacking for CSAB
9. Little psychometric development accompanies the instrument to evaluate its quality
10. Not recommended for routine assessments

Basic Skills Assessment (BSA)

Grades: 7 and over
Publication Years: 1977, 1981
4 tests: Reading (3 scores - Literal Comprehension, Inference Evaluation, Total (Forms A, B, C)
A Writer's Skills: 5 scores - Spelling, Capitalization-Punctuation, Usage, Logic-Evaluation, Total (Forms A, B, C)
Mathematics: 3 scores - Computation, Applications, Total (Forms A, B, C)
Writing Sample: A Direct Measure of Writing (Form A)

Price Data:
$26.25 per 35 Reading tests; $24.50 per 35 A Writer's Skills tests; $24.50 per 35 Mathematics tests; $17.85 per 35 Writing Sample: A Direct Measure of Writing; $3.75 per directions for administering; $3.75 per manual for scoring the writing sample
Publisher: CTB/McGraw-Hill
Comments: Tentative norms are for total scores, Grades 8, 9 and 12 only
Publisher recommends use of local norms
40-60 minutes

Reviewer 1:

1. BSA is a secondary-level testing program that offers schools a systematic method of certifying whether or not future high school graduates have met each school's specific standards of performance in the basic skills of reading, writing, and mathematics
2. The documentation provides clear and concise information on what are considered to be valid uses of the testing program and specifies examples of misuses of the tests
3. Reliability--Total test scores for Reading, A Writer's Skills, and Mathematics have internal consistency reliability estimates (KR-20) of .94, .94, and .92; no evidence for stability indices of reliability; no reliability or item data for the Writing Sample test
4. Validity--Strong evidence for content validity; concurrent validity was assessed by the relationship between test performance and teacher's assessments of the competence of students being tested; spotting borderline students (in teacher's judgments) and discriminating those who need remediation from those who do not is a task to which the BSA might be addressed, but no guidance is given on such discriminations; no information on the predictive validity
5. Detailed item-based information is of questionable validity and utility since no claims are made that the tests yield individual diagnostic information at the item level
6. Lack of remedial instructional information is possibly the biggest weakness in the program
7. If school systems decide they need an external assessment of competency in the three R's, the BSA deserves careful consideration
8. On the whole, the BSA appears to be well designed, well documented, and expertly developed and analyzed

Survey of Basic Skills (SBS)

Grades: Semester

Grade for Average Students	SBS Level (8 Levels)
K/2 - 1/1	20
1/2 - 2/1	21
2/2 - 3/1	22
3/2 - 4/1	23
4/2 - 6/1	34
6/2 - 8/1	35
8/2 - 10/1	36
9/1 - 12/2	37

Publication Year: 1985
6 Tests: (Forms P,Q)
Reading (6 scores)
a. Auditory Discrimination (Level 20)
b. Letters and Sounds (Levels 20, 21, 22)
c. Decoding (Level 20)
d. Listening Comprehension (Levels 20, 21)
e. Vocabulary (Levels 21, 22, 23, 34, 35, 36, 37)
f. Reading Comprehension (Levels 21, 22, 23, 34, 35, 36, 37)
Language Arts (3 scores)
a. Mechanics (Levels 21, 22, 23, 34, 35, 36, 37)
b. Usage (Levels 22, 23, 34, 35, 36, 37)
c. Spelling (Levels 22, 23, 34, 35, 36, 37)
Mathematics (4 scores)
a. Computation (Levels 21, 22, 23, 34, 35, 36, 37)
b. Concepts/Problem Solving (Levels 20, 21, 22, 23)
c. Concepts (Levels 34, 35, 36, 37)
d. Problem Solving (Levels 34, 35, 36, 37)
Reference Materials (Levels 23, 34, 35, 36, 37)
Social Studies (Levels 34, 35, 36, 37)
Science (Levels 34, 35, 36, 37)
Price Data: Available from publisher
Publisher: Science Research Associates, Inc.
Comments: Norm-referenced, standardized test with criterion-referenced features; 55 minutes

Review: None available at present

SUBJECT AREA TESTS

Mathematics

Design for Math

Grades: K-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6 and over
Publication Years: 1905, 1979
Mathematics Instruction System including Placement Tests, Pre- and Postdiagnostic Tests containing 14-30 subtests (2 forms) covering specific objectives; 7 levels
Price Data: (1983)

a. Levels A and B. Grades K-1 and 1-2: $150 per Level A-B kit including 45 placement tests, spirit master pretests (Form X) and posttest (Form Y), 45 profile cards, 45 award certificates, learning sequences and all necessary manuals
b. Levels C and D. Grades 2-3 and 3-4: $195 per Level C-D kit including 45 placement tests, spirit master pretests (Form X) and posttest (Form Y), 45 profile cards, 45 award certificates, learning sequences, and all necessary manuals
c. Levels E, F, and G. Grades 4-5, 5-6, 6 and over: $210 per Level E-G kit including 45 placement tests, spirit master pretests (Form X) and posttest (Form Y), 45 profile cards, 45 award certificates, learning sequences, and all necessary manuals
d. $3.75 per manual of any one level

Comments: Criterion referenced; no data on reliability; no norms; mastery ratios are given for each subtest; 50-60 minutes for the placement tests
Publisher: NCS Interpretive Scoring Systems

Reviewer 1:

1. No documentation linking the domain structure to the placement and diagnostic progress test items
2. No information is presented on how the items were developed from the structure, the number of items per objective/cluster/level, and the rationale underlying the item generation or selection procedures
3. No indices of decision validity or decision reliability are reported
4. No explanation is given for how the mastery ratio (5/6) or performance standard was set for each test

Reviewer 2:

1. Well organized, comprehensive system of mathematics instruction

2. Covers 176 skill areas for grades K through 6
3. All of the testing and support materials appear to be very good
4. No information concerning test development, and no technical data

DMI Mathematics Systems Instructional Objectives Inventory (DMI/MSIOI)

Grades: K.6-1.5, 1.6-2.5, 2.6-3.5, 3.6-4.5, 4.6-5.5, 5.6-6.5, 6.6-8.9
Publication Year: 1983
4 areas: Whole Numbers, Fractions/Decimals/Integers, Measurement and Geometry, Problem-Solving and Special Topics (7 levels)
Price Data: (1983)

a. Level A. Grades K.5-1.5: $27.50 per 35 tests and manual; 53-73 minutes
b. Level B. Grades 1.6-2.5: $29.75 per 35 tests and manual; 71-91 minutes
c. Level C. Grades 2.6-3.5: $44.95 per 35 tests and manual; 75-95 minutes
d. Level D. Grades 3.5-4.5: $33.25 per 35 tests and manual; 93-124 minutes
e. Level E. Grades 4.6-5.5: $26.25 per 35 tests and manual; 106-131 minutes
f. Level F. Grades 5.6-6.5: $26.25 per 35 tests and manual; 128-158 minutes
g. Level G: Grades 6.6-8.9: $26.25 per 35 tests and manual; 157-186 minutes

Comments: Criterion-referenced approach to mathematics assessment and instructions
No data on reliability
Estimated norms reports available for machine-scorable and hand-scorable editions
Publisher: CTB/McGraw-Hill
Reviews: Not available

Multilevel Academic Skills Inventory: Math Diagnostic Battery (MASI)

Grades: 1-8
Publication Year: 1982
2 major objectives: Computation, Applications
Price Data: (1985)
$69 per test kit including examiner's manual, resusable test, 12 response booklets, and 12 progress record forms
Publisher: Charles E. Merrill Publishing Company

Comments: Reading and Language Arts also available; criterion-referenced; survey and specific level tests designed to determine the teaching needs of students; no norms; standards of mastery provided; administration time not reported
Reviews: None available

The Steenburgen Diagnostic-Prescriptive Math Program

Grades: 1-3, 4-6
Publication Year: 1978
1 form consists of 2 parts (Quick Math Screening Test - Level 1 = Grades 1-3, and Level 2 = Grades 4-6), and Prescriptive Package for both levels
Price Data: (1984)
$25 for 25 tests and 25 profiles of each level
Comments: Criterion-referenced test; no data on reliability; no norms; no manual; 10-20 minutes
Publisher: Academic Therapy Publications

Reviewer 1:

1. The Prescriptive Tests (10 items per skill) will be useful to teachers because permission was granted by the author to duplicate these skills tests on as-needed basis
2. Neither evidence that mathematics program is based upon well-developed hierarchies of math computation skills nor that the tests in the program have sound psychometric properties

Reviewer 2:

1. No information as to what "criterion-referenced" means to the author or why this test fits that meaning
2. No domain is sampled
3. No psychometric or statistical evidence for anything connected with this program

Reading

Achievement Tests: Grades 1-8

Grades: 1, 2, 3, 4, 5, 6, 7, 8
Publication Year: 1980
1 form, 8 levels, 4 scores: Comprehension Skills, Vocabulary Skills, Decoding Skills, Study Skills

Price Data: (1984)
$7.12 per 10 tests and teacher's manual
Publisher: Macmillan Publishing Company, Inc.
Comments: Criterion-referenced
Focus is on major reading objectives "directly related to skills tested in the SERIES R Assessment Tests"
No data on reliability or validity
No norms; acceptable performance defined at 75% for each subtest
Administration time not reported
Reviews: None available

Corrective Reading Mastery Tests

Grades: 4-12 and adults in the corrective reading program
Publication Years: 1980, 1981
6 tests: Comprehension A: Thinking Basis; Comprehension B: Comprehension Skills (2 subtests); Comprehension C: Concept Applications (2 subtests); Decoding A: Word-Attack Basis (1 form, 14 scores); Decoding B: Decoding Strategies (2 subtests); Decoding C: Skill Applications (2 subtests)
Price Data: (1984)
Comprehension A: $17.85 per set (20 tests and manual)
Comprehension B: $36.95 per set
Comprehension C: $36.95 per set
Decoding A: $17.85 per set
Decoding B: $36.95 per set
Decoding C: $36.95 per set
Publisher: Science Research Associates, Inc.
Comments: Criterion-referenced
Comprehension A: 25-40 minutes
Comprehension B: 20-40 minutes per subtest
Comprehension C: 20-30 minutes per subtest
Decoding A: 35-40 minutes
Decoding B: 10-15 minutes per subtest
Decoding C: 10-15 minutes per subtest

Reviewer 1:

1. Tests include a sample of all of the skills practiced in the SRA decoding or comprehension program
2. Instructions for administration and scoring of tests are very clear
3. "Pass" criterion for each skill varies, so that for some skills one error is allowed, but for others no errors are allowed
4. No information about the relationships between student performance on the decoding and comprehension tests and other related reading tests

5. Number of items which measure each skill varies widely and is sometimes very small
6. Reservations are raised with regard to certain scoring procedures, the reliability of some test items, and lack of information about validity

Initial Placement Inventory (IPI)

Grades: 1-8
Publication Year: 1980
Selected levels to cover grades 1 through 8 in 3 areas: Silent Reading, Word Recognition, Oral Reading, 1 form (62 stencils)
Price Data: Available from publisher
Publisher: Macmillan Publishing Company, Inc.
Comments: Designed to place students in SERIES R Macmillan Reading Program
Criterion-referenced
No data on reliability
Administration time not reported
Reviews: None available

The Instant Words Criterion Test

Grades: Elementary students
Publication Year: 1980
1 form with 300 words with and without suffixes; no manual
Price Data (1983)
$8 per package of 40 tests
Publisher: Jamestown Publishers
Comments: Criterion-referenced test of student's sight reading abilities
Item scores only
No data on reliability
No norms
No suggested standards of mastery for item and total scores
15-20 minutes

Reviewer 1:

1. Test is criterion-referenced in the sense that all of these 300 words have to be known by any reader of English
2. No instructions are given on how to administer the suffix test
3. Repeated testing is advised; there is, however, only one form of the test, by definition
4. No distinction is made between lexical and non-lexical items

5. Although words chosen may be among the most frequently used in fluent native writings, or indeed in children's books, there is no guarantee that they are the ones most used by learners
6. No comments provided on reliability or validity
7. No concurrent or predictive data
8. The Instant Words Criterion Test cannot be easily assessed using criteria applicable to either psychometric or criterion-referenced tests other than to regard it as extremely inadequate

The Laurita-Trembly Diagnostic Word Processing Test (DWPT)

Grades: 1.9-4.9, 5.0-College
Publication Year: 1979
1 form, 2 levels (test materials exactly same for both levels, only different directions); 4 scores: Raw, Word Correct, Vertical Word Processing, Horizontal Word Processing derived from 6 areas: Vowel Discrimination, Direct Phonic Processing, Indirect Structural Processing (long vowel, short vowel), Prefixes and Suffixes (long vowel, short vowel)
Price Data: (1982)
$24.95 per set of testing materials
Publisher: L & T Educational Materials, Inc.
Comments: Criterion-referenced
No data on reliability and validity
No norms
Administration time not reported

Reviewer 1:

1. Test is designed primarily according to the strong beliefs of the authors, not according to evidence
2. Test purports to diagnose learning potential, to direct the design of instruction to a mastery criterion, and to evaluate the effectiveness of this instruction. However, in actuality it does not and in fact cannot do these things due to logical flaws in the design of the program
3. The domains sampled by the test are ill defined. This renders criterion-referenced interpretation of such scores as the word count difficult at best
4. With respect to the essential topic of content validity, there is no explanation of the method or rationale of the word sampling
5. Unable to recommend it for any applied purpose

The Macmillan Reader Placement Test

Grades: 1 and over
Publication Years: 1967, 1972
2 tests: Word Pronunciation, Oral Reading; Forms 1 and 2
Price Data: Available from publisher
Publisher: Macmillan Publishing Company, Inc.
Comments: Criterion-referenced
Part of Macmillan Mastery Tests
No data on reliability
Basis for score interpretation not presented
20 (25) minutes administration time
Reviews: None available

Mastery: An Evaluation Tool: Selected Short SOBAR Reading Tests (SOBAR-Systems for Objective-Based Assessment-Reading)

Grades: 3-9
Publication Years: 1975, 1976
9 tests with varying grade levels: Phonic Analysis (grades 3-5); Structural Analysis (grades 3-9); Study Skills (Grades 4-9); Vocabulary I (grades 4-5); Vocabulary II (grades 6-7); Vocabulary III (grades 8-9); Comprehension I (grades 4-5); Comprehension II (grades 6-7); Comprehension III (grades 8-9)
Price Data: (1983)
$23.45 per 25 tests, 1 examiner's manual, and 1 examiner's supplement (specify test)
Publisher: Science Research Associates, Inc.
Comments: "Provides an alternative to the full-length catalog Mastery Test in SOBAR Reading"
Criterion-referenced
No data on reliability or validity
No norms
60-75 minutes per test

Reviewer 1:

1. SOBAR Short Catalog Tests are part of a larger SRA evaluation program called Mastery: An Evaluation Tool. The Mastery builds objective-based, criterion-referenced tests for math and reading for grades 3 through 9. There are 138 objectives grouped into 5 major categories
2. This shorter version of the grade-based tests reportedly was developed in response to teacher requests for a faster, category-specific evaluation program
3. There is no explanation for how the objectives chosen from each category were determined

4. No item difficulty information; difficult to judge whether failure on an item is due to lack of mastery of an objective or to the difficulty of the item
5. The lack of a conceptual framework, the absence of technical information (item reliability, validity, discrimination, relevance, and difficulty), the complexity and arbitrariness of the items, and the availability of plausible alternatives all argue against the use of this system

McLeod Phonic Worksheets

Grades: Reading level grade 3.5 and under
Publication Years: 1972, 1977
6 areas: Initial Single Consonants, Final Single Consonants, Single Vowels, Consonant Blends and Digraphs, Vowel Blends, Auditory Discrimination
Price Data: (1984)
$22 per complete kit consisting of 4 test cards, 5 record and analysis masters, 63 worksheet masters, and teacher's guide
Publisher: Educators Publishing Service, Inc.
Comments: Formerly called Domain Phonic Tests
Criterion-referenced
No data on reliability and validity
No norms
Individual test
No specific manual; administrative, descriptive and interpretive information included in same booklet as tests and worksheets
Administration time not reported
Reviews: None available

Multilevel Academic Skills Inventory: Reading and Language Arts (MASI)

Grades: 1-8
Publication Year: 1982
1 form, 5 major objectives: Decoding, Comprehension, Content Vocabulary, Handwriting, Spelling
Price Data: (1985)
$69 per test kit including examiner's manual, reusable test, 12 response booklets, and 12 progress record forms
Publisher: Charles E. Merrill Publishing Company
Comments: Math Diagnostic Battery also available
Criterion-referenced
Survey and specific level tests designed to determine the teaching needs of students
No norms

Standards of mastery provided
Individual test
Administration time varies according to test
Reviews: None available

PRI Reading Systems (PRI/RS)

Grades: K-1, 1-2, 2-3, 4-6, 7-9 (5 levels)
Publication Year: 1980
System 1: A Graded Approach (2 tests per level: Instructional Category)
System 2: A Multi-Graded Approach (2 tests: Placement Test and Skill Area Diagnostic Test)
Both systems with 4 basic skill areas: Oral Language, Word Attack and Usage, Comprehension, Applications
Price Data: (1984)
System 1 = $9.80 per 35 locator tests for Level A-C or C-E; $9 per 50 answer sheets; $4.25 per examiner's manual
System 2 = Placement Test: $12.95 per 35 hand-scorable or reusable tests in one area; $7 per 50 answer sheets for any one area
= Skill Area Diagnostic Test: $12.60 per 20 tests of any one area; $5.95 per 35 hand-scorable answer sheets; $4.25 per examiner's manual for any one area
Publisher: CTB/McGraw-Hill

Reviewer 1:

1. This program is surely one of the best among criterion-referenced programs in the comprehensiveness of its tests
2. Despite a broad conception of reading, however, this assessment program suffers from the major deficiencies that have typically characterized criterion-referenced programs which are not inherent to them
3. Decisions on students' mastery of objectives are made on performance on a small number of items for which the reliability indices are, on the average, moderate
4. If this program is intended to assess knowledge of an entire reading curriculum, some serious questions can be raised about the comprehensiveness of the objectives
5. Because of the failure to provide information that children actually require these skills for proficient reading and that children learn in the outlined sequence, teachers need to be cautioned in their use of the test results
6. Face validity of the test is generally good. Instructions, sample items, test sheets, and information about timing all contribute to this

7. A program such as this can be used as one component of a classroom reading program
8. Teachers should be cautioned to use results from other measures and from their observations as they interpret data from these tests

Reading Yardsticks (RY)

Grades: K, 1, 2, 3, 4, 5, 6, 7, 8
Publication Year: 1981
1 form, 9 levels: Level 6 (Grade K) to Level 14 (Grade 8)
Price Data: (1984)
$.84 per directions for administration; $1.62 per answer key for all levels; $3.63 per teacher's guide; $6.60 per technical report
Publisher: Riverside Publishing Company
Comments: Criterion-referenced
Optional eyesight for Grades 4-8

Reviewer 1

1. All levels of items match logically and directly with the stated objectives for the level
2. Reported reliabilities of differences for the Reading Yardstocks, though not extremely high, are impressive in the context of previous research in the area
3. Correlations between RY and IYBS subtests, which are generally in the .60s and .70s, are fairly strong evidence of concurrent validity
4. Directions for administering tests are clear and precise
5. Use of test results is not as well presented as would be desirable
6. Major limitation for the potential user of RY is that an alternative form of the test is not available
7. Test should have significant appeal to reading teachers whose philosophical approach to reading fits the belief that general reading skill is composed of subskills that can be taught independently of one another
8. Publisher has failed to make specific recommendations, based on the available data, to make the programs easy to handle for someone without substantial experience working with testing program

Wisconsin Tests of Reading Skill Development: Comprehension (WTRSD:C)

Grades: K-2, 1-3, 2-5, 3-6, 4-6, 5-6, 6
Publication Year: 1977

7 levels, 2 formats for Levels A-D (Consumable Booklet, Ditto Masters)
Price Data: (1984)
$21 per set of profile cards; $2.25 per administrator's manual for any one level; $5 per teacher's planning guide
Publisher: NCS Interpretive Scoring Systems
Comments: Part of the Wisconsin Design for Reading Skill Development, which also includes word attack and study skills assessment
Based on a 1974 developmental edition
Criterion-referenced
No norms

Reviewer 1

1. The Comprehension assessment component is an objective-based testing program which corresponds to the skills/objectives set out in the Design approach
2. The various forms and format provided flexibility in use and opportunity for retesting without using the same test items for a particular objective
3. Because the number of items per score is usually between 10 to 20, the reliability coefficients should be higher than for some other criterion-referenced measures which spread the items thinly over many objectives
4. No discussion of how these reliabilities (the .70s and .80s) were calculated and no supporting documentation was provided
5. No information relating to either item-test or item-criterion relationships
6. Except for general information on content validity, presumed educational importance, and to a slight extent construct validity, no information on validity is included in the manual
7. Many of the tests appear to be soundly developed to meet their expressed purposes. One exception is the word parts substrand of the Word Meaning tests
8. A potential problem with passage dependency was also noted with the central thought items of the Passage Meaning tests, although gender balance seems admirable

Spelling

Spellmaster

Grades: 1 to adult
Publication Years: 1974, 1976
4 tests: 8 Irregular Word Tests, 8 Homonyms Tests, 8 Diagnostic Tests with 8 overlapping levels depending on individual's ability

Price Data: (1986)
$9.95 per 30 tests of any one diagnostic test; $6.95 per Spelling = Testing and Evaluating; $8.95 per Spelling = Teaching and Learning
Publisher: Learnco Inc.
Comments: Criterion-referenced tests; no norms; suggested standards of mastery; 20-30 minutes per test

Reviewer 1

1. Primarily designed for elementary and junior high school students
2. Each of the Spellmaster tests examines recall processes and requires the testee to write words that have been dictated and used in an oral sentence
3. Without adequate information about validity and reliability, decisions based on results must be tentative
4. The relationship among the test items, the domain of the test content, and the methods of measuring that content were not adequately documented
5. The Diagnostic Tests do not appear to be a valid measure of phonemic skills
6. Spellmaster cannot be recommended for use as global measure of spelling ability, as an indication of spelling mastery at a particular level, or as a determinant of the ability to spell specific elements within words

Composition

Writing Proficiency Program (WPP)

Grades: 6-9, 9-10, 11-13
Publication Years: 1979, 1981
3 levels: Intermediate System (Grades 6-9); Test 1 (Grades 9-10); Test 2 (Grades 11-13); 2 parts each level (Multiple Choice, Writing Exercises)
Price Data:
$23.45 per 35 multiple-choice tests; $11.50 per 50 of any one writing exercise
Publisher: CTB/McGraw-Hill
Comments: Criterion-referenced assessment and instructional system
No data on reliability or validity
Mastery criterion: 80% for each test objective
30-50 minutes per part per level

Reviewer 1

1. Tests consist of timed, criterion-referenced, multiple-choice sections as well as a spontaneous writing sample that varies dependent on the grade level of the examinee
2. No technical characteristics; no reliability and validity
3. Multiple-choice subtests have face validity although this certainly is not a sufficient basis for confidence concerning the content validity of the tests
4. Scoring instructions for the spontaneous written language portion of the tests are included in the separate booklets entitled "Guide to Writing Samples"
5. Tests that accompany the WPP have merit primarily for usage within the overall instructional program
6. Multiple-choice tests, when evaluated separately from the overall program, do not offer any significant features that set them apart from other available instruments
7. Writing samples and accompanying procedures are noteworthy for their degree of detail and rationale but again are useful mainly within the context of the overall program

APPENDIX F

NORTH CENTRAL ACCREDITATION PROCESS

NORTH CENTRAL ACCREDITATION PROCESS

Since North Central Association membership is a voluntary activity and affects a limited number of Nebraska schools, state accreditation and evaluation should be separate and distinct from North Central activities.

The North Central Association has developed a process for outcomes accreditation and evaluation and has listed some indictors of effective school practices.

Indicators of Effective School Practices

- Mission and goals statement
- Data gathering for decision making
- Appropriate expectations
- Emphasis on learning; a comprehensive curriculum
- School monitors teaching/learning
- Climate
- Leadership moves school toward goals
- School/home partnership programs
- Clear relationships--curriculum objectives, practice

The Association lists external contextual data that should be considered in evaluating school performance.

External Contextual Data

- Free and reduced lunch students
- Formal education of parents
- Predominant community approach to child rearing
- Per capital income
- Cognitive abilities of children in preschool or kindergarten
- Percent minority
- Level of staff preparation
- Per pupil expenditure

Input Indicators

Standard II	Organization, Administration and Control Staffing adequacy--custodial, clerical, full-time administrators required
Standard III	Instructional Program Carnegie units (38) provided; courses provided--language arts, etc.

Standard IV	Professional Staff Professional preparation--hours per subject, staff/student ratios, teacher load
Standard V	Student Activities Provide qualified sponsors Provide opportunities, balance
Standard VI	Student Personnel Services Health--nurse available Counselors--ratio 1/450 Provide for special need facilities
Standard VII	School Improvement Activities Provide sufficient autonomy, resources
Standard VIII	Instructional Media Program Adequacy of collection, minimum standards Number of stations based on enrollment Number of professional staff provided Expenditure per student
Standard IX	Financial Support and Control Provide sufficient funds to meet standards
Standard X	School Facilities, Instructional Materials Provide adequate space, facilities Provide adequate equipment
Standard XI	Evaluation
Standard XII	Adult Education
Standard XIII	School/Community Relations Provide planned program

Process Indicators

Standard I	Pholosophy and Goals (written)
Standard II	Organization Administration and Control Administrative services Reports and records Board/staff relationships Board rules and policies Working relationships School day and year Graduation requirements

Standard III	Instructional Program Quality programs accessible, climate, classroom practices, assessment, communications, articulation, community resources, special education, instructional objectives
Standard IV	Professional Staff Preparation, transcripts, in-service
Standard V	Activities Balance, opportunity
Standard VI	Personnel Services Health--inventory and diagnosis of special needs, health records Guidance--provide services
Standard VII	School Improvement Activities Promote managed change Involvement of staff
Standard VIII	Instructional Media Program Provide services Maintain records, materials
Standard IX	Financial Support and Control Budgetary procedures Equitable distribution of resources
Standard X	School Facilities, Instructional Materials Accessible site Safety precautions met
Standard XI	Evaluation Seven-year self-study Processes for evaluation Continuous evaluation, follow-up study
Standard XII	Adult Education
Standard XIII	School Community Relationships Provide for community involvement Provide communications activities

APPENDIX G

SELECTED REFERENCES

SELECTED REFERENCES

Archbald, D. A. and Newman, F. M. *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School.* Reston, Virginia: National Association of Secondary School Principals, 1988.

District 66, Westside Community Schools. *Elementary School Time Allocations.* Omaha: District 66, Westside Community Schools.

Educational Service Unit #3. *Textbook Information for Elementary Schools.* Omaha: Educational Service Unit #3, Cass, Douglas, Sarpy, and Washington Counties.

Illinois State Board of Education. *The State's Role in Regulating Education: Report and Preliminary Recommendations.* Springfield, Illinois: State Board of Education, 1987.

Illinois State Board of Education. *The Illinois Program for Evaluation, Supervision, and Recognition of Schools* (10M 6-452B-38, No. 552). Springfield, Illinois: State Board of Education, 1986.

Illinois. *An Act in Relation to Educational Reform and the Financing Thereof* (Revised Statutes, P.A. 84-126), 1985.

Illinois. *The School Code,* 1988.

Illinois State Department of Education, Department of School Improvement Services. *State Goals for Learning and Sample Learning Objectives: Language Arts; Mathematics; Biological and Physical Sciences; Social Sciences; Physical Development and Health; and Fine Arts* (10M 55923). Springfield, Illlinois: State of Illinois, 1986.

Medina, Noe and D. Monty Neill. *Fallout from the Testing Explosion.* Cambridge, Massachusetts: National Center for Fair and Open Testing, 1988.

National Association of Secondary School Principals. *Education Indicators: A National Report Card.* Reston, Virginia: National Association of Secondary School Principals.

National Governor's Association. *Governors' 1991 Report on Education; Results in Education: 1988.* Washington, D.C.: National Governors' Association.

Nebraska Department of Education, Approval and Accreditation. *The Nebraska Model: A Model for Comprehensive School Evaluation.* Lincoln: The Department, 1988.

Nebraska Department of Education, Approval and Accreditation. *Planning the Weekly Schedule in the Elementary Grades: Planning the Use of Time in Kindergarten.* Lincoln, Nebraska: Nebraska Department of Education, 1988.

Nebraska Department of Education, School Management Services. *Comprehensive School Evaluation: Organizing the External Visitation.* Lincoln: The Department.

Nebraska Department of Education, School Management Services. *Comprehensive School Evaluation: The Role of the External Evaluator.* Lincoln: The Department.

Nebraska Department of Education, School Management Services. *Comprehensive School Evaluation: Implementation of Recommendations.* Lincoln: The Department.

Nebraska Department of Education, School Management Services. *Comprehensive School Evaluation: Organizing the Self-Study.* Lincoln: The Department.

Nebraska Department of Education, Management Information Services. *Instructional Manual for the Fall Personnel, Curriculum, School District Membership, and State Aid Supplement.* Lincoln: The Department, 1988.

Nebraska Department of Education, Management Information Services. *Report of 1987-1988 Counts of Class 2-6 School Districts That Offer/Do Not Offer Courses in Grades 7-12, Excluding SPED, Self-Contained and Unclassified Courses* (NDE 02-015 Fall Curriculum Report). Lincoln: The Department, 1988.

North Carolina Department of Public Instruction. *Guide to Test Development.* Raleigh: The Department, 1988.

North Central Association of Colleges and Schools. *The NCA OA/E Guide: A Handbook for Outcomes Accreditation/Evaluation Schools.* Boulder, Colorado: Commission on Schools, 1988.

North Central Association of Colleges and Schools. *Policies and Standards for the Accreditation of Optional Schools and Special Function Schools; Standards for Secondary Schools; Standards for Elementary Schools; Standards for Middle Level Schools; and Standards for Independent College-Preparatory Schools.* Boulder, Colorado: Commission on Schools, 1988.

North Central Association of Colleges and Schools. *Documenting Student Gains in the Outcomes Accreditation School: Guidelines to Be Used in Conjunction with the NCA OA/E Guide: A Handbook for Outcomes/Accreditation Schools.* Boulder, Colorado: Commission on Schools, 1988.

North Central Association of Colleges and Schools. *Partners in Quality Education: The Community, Its Schools, and the North Central Association.* Boulder, Colorado: Commission on Schools, 1988.

O'Connell, P. A. *An Analysis of Selected 1985-1986 School Report Card Variables and Student and District Performance in Illinois Public Schools.* Unpublished Doctoral Dissertation, Illinois State University, 1987.

Office of Educational Research and Improvement, U.S. Department of Education, Center for Education Statistics. *The Condition of Education: A Statistical Report* (CS 87-365). Washington, D.C.: Government Printing Office, 1987.

Office of Educational Research and Improvement, U.S. Department of Education, Center for Education Statistics. *A Classification of Instructional Programs* (CIP) (CS 87-308). Washington, D.C.: Government Printing Office, 1987.

Omaha Public Schools. *Textbook Information for Elementary Schools.* Omaha: Omaha Public Schools.

Programs for the Improvement of Practice, Office of Educational Research and Improvement, U.S. Department of Education. *Creating Responsible and Responsive Accountability Systems: Report of the OERI State Accountability Study Group* (PIP 88-808). Washington, D.C.: U.S. Government Printing Office, 1988.

South Carolina State Department of Education. *Guidelines for the Administration, Evaluation, and Accreditation of South Carolina Schools and Districts.* Charleston: The Department, 1987.

State of Nebraska. *School Laws.* Fremont, Nebraska: Hammond and Stephens, 1988.

United States Department of Education, Office of Planning, Budget and Evaluation. *Wall Charts: State Performance Outcomes, Resource Inputs, and Population Characteristics, 1982 and 1984; Student Performance, Resource Inputs and Population Characteristics, 1982 and 1985; Student Performance, Resource Inputs, State Reforms, and Population Characteristics, 1982 and 1986; Student Performance and Resource Inputs, 1985 and 1986.* Washington, D.C.: U.S. Government Printing Office.